The role of miRNA regulation in cancer progression and drug resistance

Joshi, Tejal

Publication date:
2012

Document Version
Publisher's PDF, also known as Version of record

The role of miRNA regulation in cancer progression and drug resistance

Tejal Joshi

29^{th} February, 2012

CENTERFO
RBIOLOGI
CALSEQU
ENCEANA
LYSIS CBS

Contents

Preface

This thesis was prepared at the Center for Biological Sequence Analysis (CBS), Department of Systems Biology, at the Technical University of Denmark (DTU) in partial fulfilment of the Ph.D. degree. The Ph.D. was funded by a DTU scholarship, Tamoxifen resistance project was in collaboration with the Sino Danish breast cancer research center (funded by Danmarks Grundforskningsfond) and the MALT/DLBCL project was in collaboration with Eye pathology unit of Rigshospitalet, University of Copenhagen, Denmark.

All the work was carried out at the Center for Biological Sequence Analysis under the supervision of Associate Professor Christopher Workman and Associate Professor Ramneek Gupta.

Lyngby, February 2012

Tejal Joshi

Abstract

This PhD thesis presents the work carried out at Center for Biological Sequence Analysis, Technical University of Denmark. The projects presented in this thesis are a purely bioinformatic in nature. Included in this thesis are the two projects that focus on the gene regulatory events mediated by miRNAs in the context of cancer biology, drug resistance and disease progression.

The first project described in Chapter 6 addresses the problem of tamoxifen resistance, an anti-estrogen drug that is generally highly effective in the treatment of ER-positive breast cancers. The underlying molecular mechanisms for the acquired resistance to tamoxifen are not very well understood. Therefore, with the aid of miRNA and gene expression profiles for MCF7/S0.5 (tamoxifen sensitive) and three MCF7/S0.5 derived tamoxifen resistant cell lines, we obtained several miRNA-mediated regulatory events in the tamoxifen resistant cell lines. Following a systems biology approach of integrating evidences of functional interactions such as transcription factor (TF)-miRNA interactions, we have identified a number of biologically relevant pathways involved in the development of tamoxifen resistance.

Chapter 7 presents a study highlighting the role of miRNAs in the transformation of ocular mucosa associated lymphoid tissue lymphoma (MALT) to the high-grade diffuse large B-cell lymphoma (DLBCL) of eye. Several tumor suppressive miRNAs were found to be dysregulated in DLBCL, suggesting their possible role in disease transformation. Many of those were under transcriptional regulation by MYC and NFKB1, the key transcription factors involved in lymphomas. Furthermore, upstream regulators of NFKB1 were also repressed, suggesting a possible loss of regulation of NFKB1 may contribute to the activation of NF-κB signaling pathway, and thereby to the disease transformation.

In summary, this thesis focuses on regulatory role of miRNAs in drug resistance and disease progression. The findings provide hints toward various biologically and perhaps therapeutically relevant gene regulatory events. This thesis demonstrates the right choice of data analysis techniques combined with a systems biology approach provides better understanding of the complex biology.

Dansk resumé

Denne ph.d.-afhandling præsenterer arbejde, der er udført på Center for Biologisk Sekvensanalyse, Danmarks Tekniske Universitet. De projekter, der præsenteres i denne afhandling er rent bioinformatiske af natur. Inkluderet i denne afhandling er to projekter, der fokuserer på de gen-regulerende begivenheder medieret af miRNA i forbindelse med cancer biologi, lægemiddelresistens og sygdomsprogression.

Det første projekt beskrevet i kapitel 6 addresserer problemet vedrørende tamoxifen resistens, et anti-østrogen stof, som generelt er yderst effektiv i behandlingen af ER-positive brystcancere. De underliggende molekylære mekanismer i erhvervet resistens over for tamoxifen, er ikke særlig godt forstået. Derfor, ved hjælp af miRNA- og genekspressionsprofiler for MCF7/S0.5 (tamoxifen følsom) og tre MCF7/S0.5 afledte tamoxifen resistente cellelinier, kom vi over adskillige miRNA-medierede regulatoriske begivenheder i tamoxifen resistente cellelinier. Efter en systembiologisk tilgang til at integrere beviser for funktionelle interaktioner såsom transskription faktor (TF)-miRNA interaktioner, har vi identificeret en række biologisk relevante veje, der er involveret i udviklingen af tamoxifen resistens.

kapitel 7 præsenterer en undersøgelse der fremhæver miRNAers rolle i omdannelsen af okulære slimhinde associeret lymfevæv lymfom (MALT) til den højt rangeret diffuse storcellet B-celle lymfom (DLBCL) i øjet. Flere tumor undertrykkende miRNA viste sig at være fejlreguleret i DLBCL, hvilket tyder på deres mulige rolle i sygdom transformation. Mange af dem var under transkriptionel regulering af MYC og NFKB1, de vigtigste transkriptionsfaktorer involveret i lymfomer. Desuden var opstrøms regulatorer af NFKB1 også undertrykt, hvilket tyder på et muligt tab af reguleringen af NFKB1 kan bidrage til aktivering af NF-κB signalvejen, og dermed aktiveringen af sygdomstransformation.

Sammenfattet fokuserer denne afhandling på den regulatoriske rolle af miRNA i lægemiddelresistens og sygdomsprogression. Resultaterne giver antydninger af forskellige biologiske og måske terapeutiske relevante genregulatoriske begivenheder. Denne afhandling viser det rigtige valg af dataanalyseteknikker kombineret med en systembiologi tilgang giver bedre forståelse af den komplekse biologi.

Acknowledgements

It has been a great pleasure to be a PhD student at Center for Biological Sequence analysis. My deep gratitude toward the center director Professor Søren Brunak, for creating such a friendly and inspiring work environment. Everyone here at CBS has always been willing to help when needed. I have always appreciated the competitiveness of the colleagues and the warmth in the work environment here at CBS.

My special thanks to my supervisor Chris Workman. I have learnt a lot from you. You have always put emphasis on the systematic data analyses, better data visualization and statistics; I wish I could learn even more from you. I will always be grateful for all the support and thoughtfulness you have shown during my stay in Denmark and at CBS.

Many thanks to Ramneek Gupta, my CBS collaborator. You deserve a big part of this thesis coming through. Thank you for believing in my scientific abililities and making me a part of the SinoDanish group. I am also thankful to Jan Stenvang, Henrik Ditzel, Nils Brunner, Daniel Elias, Peter K. Rasmussen, Kirsten Gronbeck, Christoffer Hother and Louise Borst for broadening my scientific horizons.

Thanks to Zoltan, Aron, Laurent, Simon and Kirstine for the stimulating scientific discussions and the help.

My very special gratitude to my mom, dad and my brother, who have been very patient during these three years and have always been there for me when I needed. This PhD would not have been possible without you.

Special thanks to Agata, Juliet and Greg for being such great friends. I have met many nice and kind people at CBS with whom I have enjoyed having a quick conversation in the CBS kitchen or having a lunch together. Thanks to Josef, Bent, Salvatore, Pia, Daniel (Aaen), Henrik (Nielsen), Ali, Rachita, Oksana, Edita, Natasja, Erwin, Aline, Sophie, Sara, Fred, Nils, Tune, Hanne, Dhanny, Kasper, Arcadio, Irene, Ida, Manos, Stranzl.

Thanks to the excellent IT support personnel, John, Kristoffer, Peter Wad, Olga for the technical support. CBS office administration deserves a big thanks. Dealing with the bureaucratic aspects associated with an international employee always requires little more work; thanks to Lone, Dorthe, Marlene, Annette and Louise for your help.

Thanks to Martin for your help in understanding complicated aspects of statistical learning and also for just being there to motivate me when I needed it the most.

And lastly to Thomas Arn, who have been very patient and extremely helpful. Thanks for the cheerful mood and encouraging words.

Papers included in the thesis

- **Tejal Joshi**, Daniel Elias, Jan Stenvang, Maria Lyng, Anne Lykkesfeldt, Nils Brünner, Ramneek Gupta, Christopher T. Workman, Henrik J. Ditzel. *Integrative analysis of miRNA and gene expression reveals regulatory networks in tamoxifen resistance.* Manuscript in preparation.

- Christoffer Hother*, Peter K Rasmussen*, **Tejal Joshi**, Ditte Reker, Ulrik Ralfkiær, Christopher T. Workman, Steffen Hedegaard, Elisabeth Ralfkiær, Kirsten Grønbæk. *MicroRNA profiling suggest a role for MYC and NFKB1 mediated dysregulation of microRNA expression in the transformation of low-grade ocular MALT to diffuse large B-cell lymphoma.* Manuscript in preparation.

* These authors contributed equally.

Papers not included in the thesis

- L Borst*, A Wesolowska*, **T Joshi**, R Borup, FC Nielsen, MK Andersen, OG Jonsson, PS Wehner, F Wesenberg, BMF Frost, R Gupta and K Schmiegelow. *Genome-wide analysis of cytogenetic aberrations in ETV6/RUNX1-positive childhood acute lymphoblastic leukaemia.* In Press. *British Journal of Haematology.* Feb 8, 2012.

- S Will, **T Joshi**, IL Hofacker, PF Stadler R Backofen. *LocARNA-P: Accurate Boundary Prediction and Improved Detection of Structural RNAs.* In Press. *RNA*, January 2012.

- H Polur, **T Joshi**, CT Workman, G Lavekar, I Kouskoumvekaki. *Back to the Roots: Prediction of Biologically Active Natural Products from Ayurveda Traditional Medicine. Molecular Informatics.* Volume: 30, Issue: 2-3, Pages: 181-187, 2011.

- TA Hansen*, A Fazio*, **T Joshi**, H Polur, L Gautier, CT Workman. *The role of checkpoint kinases MEC1 and RAD53 in the cell cycle checkpoint response in S. cerevisiae.* Manuscript in preparation.

* These authors contributed equally.

Abbreviations

DNA	Deoxy Ribonucleic Acid
cDNA	Complementary DNA
IgG	Immunoglobulin G
IHC	Immunohistochemistry
ISH	in situ hybridization
LH	Luteinizing hormone
LHRH	Luteinizing-hormone-releasing hormone
MAPQ	Mapping quality
RNA	Ribonucleic Acid
miRNA	microRNA
miR	miRNA
mRNA	Messenger RNA
ER	Estrogen Receptor
ER+	Estrogen Receptor positive
TAMR	Tamoxifen resistance
logFC	Log (base 2) fold change
Adj.p-value	Adjusted p-value
FDR	False Discovery Rate
qPCR	quantitative Polymerase Chain Reaction
PCA	Principal component analysis
SVD	Singular value decomposition
UTR	Untranslated region
RT-PCR	Reverse transcription polymerase chain reaction
RNA-seq	RNA sequencing
SERD	Selective estrogen receptor downregulator
SERM	Selective estrogen receptor modulator
PCC	Pearsons Correlation Coefficient
eMALT	Mucosa Associated Lymphoid Tissue Lymphoma of eye
eDLBCL	Diffuse Large B-Cell Lymphoma of eye
C_p	Crossing point
C_t	Threshold cycle
qPCR	Quantitative Real Time Polymerase Chain Reaction
TF	Transcription factor
FI	Functional interaction
DLBCL	Diffuse large B-cell lymphoma
MALT	Mucosa associated lymphoid tissue lymphoma
WHO	World health organisation
PC	Principal component
DE	Differential expression
RISC	RNA induced silencing complex
Ago	Argonaute protein
Tm	Melting temperature
FOXM1	Forkhead box protein M1
MYC	v-myc myelocytomatosis viral oncogene homolog (avian)
YWHAG	14-3-3 gamma
YWHAZ	14-3-3 zeta

Part I

Introduction

Chapter 1

MicroRNAs

MicroRNAs (miRNAs) are 19-22 nucleotide long small noncoding RNAs that regulate expression of several genes via post-transcription repression or degradation in both animals and plants. Due to their huge regulatory potential, miRNAs have garnered a great amount of research interest. This chapter outlines the current understanding of miRNA biogenesis and moves further to highlight the mechanisms of miRNA-mediated gene regulation, target prediction algorithms and roles of miRNAs in cancer.

1.1 miRNA biogenesis

After the discovery of $lin-4$ and $let-7$ in *Caenorhabditis elegans* [108, 148], miRNAs were reported in mammals [103, 107], plants and viruses. miRNAs stem from more than 1kb long primary (pri-) miRNAs. The pri-miRNAs are synthesized by RNA polymerase II (Pol II). In mammals, a complex involving Drosha and Pasha proteins form a microprocessor complex with pri-miRNAs. This cleavage of pri-miRNA results in the double stranded miRNA precursors (pre-miRNAs). Pre-miRNAs are hairpin loops of about 150 nt lengths. As shown in Figure 1.1, the pre-miRNAs are exported from nucleus to the cytoplasm by the Exportin, where Dicer, an RNAase enzyme and its binding partner TRBP/Loqs separate two strands and form an unstable miRNA:miRNA* duplex of 22 nt length. One of the strands from this duplex, called the guide strand, is selected by Argonaute and loaded into the RNA-induced Silencing Complex (RISC) where miRNA binds with its mRNA targets [103, 72, 96, 106]. The opposite strand, now a passenger strand, is normally degraded, but recent studies using deep sequencing technologies have made it clear that the miRNA* strand often remains viable

and coexists with the mature miRNA from the opposite arm.

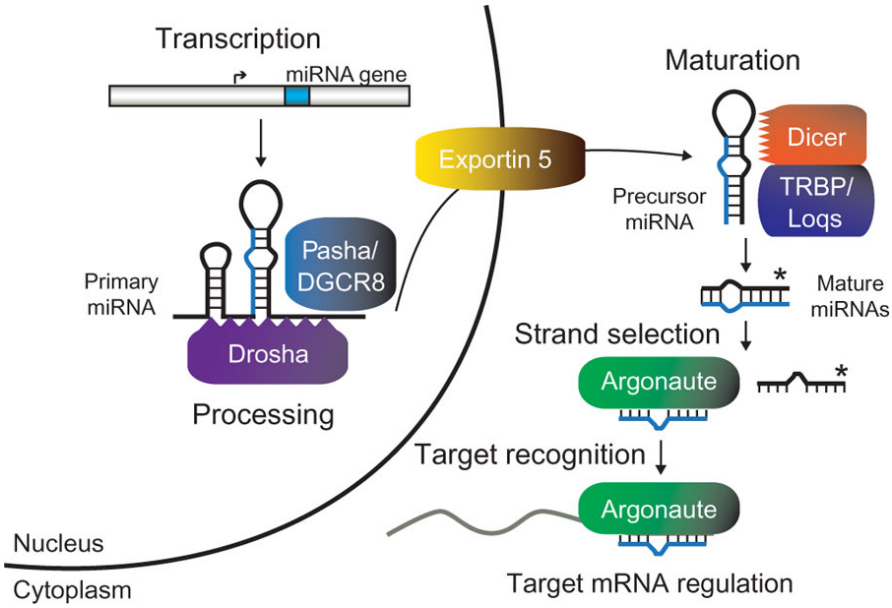

Figure 1.1. MicroRNA biogenesis [92]

MicroRNAs typically transcribe from a variety of genomic regions such as, intergenic regions, introns of protein coding genes or intron/exons of protein and/or noncoding genes. At least 50% of human miRNAs are intronic, meaning that precursor miRNAs are transcribed from the introns of protein coding or nonprotein-coding genes [154], whereas about 10% are encoded from the exons of protein-coding or nonprotein-coding transcripts. This suggests that the expression of miRNA is often related to the expression of the host transcripts from which they transcribe. Many intronic miRNAs lack their own promoter, and utilize the promoter of the gene they reside in. Therefore, the posttranscriptional regulation of the host genes often affects miRNA expression. Chien et al [31] recently published miRStart database presenting a compiled information on miRNA and transcription start sites (TSS) based on experimental evidences, covering 442 intronic miRNA-host pairs in human genome. Using the miRNA and mRNA expression data described in Chapter 6, Paper I, I found that 37 of the 156 miRNA-host gene pairs considered showed high correlation of their expression levels measured across breast cancer cell lines. In Table 1.1, the negative value of Pearson's correlation coefficient (PCC) represent an agreement of miRNA expression (measured by qPCR method) and mRNA expression (microarray) for each miRNA-host pair considered. miRNA-host pairs showing positive correlation

(PCC<0, due to qPCR vs RMA values) might be co-transcribed. It is also commonly believed that the hosted intronic miRNAs exert functions that complement the functions of the host genes they are transcribed from.

Table 1.1. Intronic miRNA-host gene pairs with high degree of correlation of expression

miRNA	**Host gene**	**PCC**	**miRNA**	**Host gene**	**PCC**
miR-107	PANK1	-0.68	miR-30e	NFYC	-0.63
miR-10b	HOXD3	-0.9	miR-335	MEST	-0.96
miR-1182	FAM89A	-1	miR-378	PPARGC1B	-0.6
miR-1201	CCNB1IP1	-0.89	miR-500	CLCN5	-0.97
miR-1243	ANK2	-0.99	miR-504	FGF13	-0.96
miR-1249	C22orf9	-0.9	miR-564	TMEM42	-0.82
miR-1258	ZNF385B	-1	miR-586	SUPT3H	-0.96
miR-1266	MYO5C	-0.87	miR-589	FBXL18	-0.93
miR-1296	JMJD1C	-0.77	miR-600	STRBP	-0.85
miR-147b	SPATA5L1	-0.72	miR-604	SVIL	-0.91
miR-149	GPC1	-0.89	miR-620	MED13L	-0.91
miR-152	COPZ2	-0.97	miR-639	TECR	-0.67
miR-15b	IFT80	-0.73	miR-640	GATAD2A	-0.88
miR-1908	FADS1	-0.84	miR-643	ZNF766	-0.96
miR-1912	HTR2C	-0.6	miR-744	MAP2K4	-0.91
miR-1914	UCKL1	-0.77	miR-941	DNAJC5	-0.8
miR-22	C17orf91	-0.88	miR-95	ABLIM2	-0.81
miR-23b	C9orf3	-0.64	miR-98	HUWE1	-0.85
miR-27b	C9orf3	-0.73			

miRBase, an online repository of miRNA sequences and their annotation, is a widely accepted source of miRNA information. The database was established in 2002 to provide a consistent nomenclature for novel miRNAs along with primary evidence supporting their annotations. Guidelines for the miRNA nomenclature were defined in 2003 [3]. In order for a 22 nt sequence to receive a miRNA status, it must fold into a stable hairpin structure characterized by low minimum free energy (MFE). Furthermore, the putative miRNA must have derived from, either an experimental cloning, homology to the known miRNAs (with or without experimental verification) or miRNAs captured by sequencing technologies. Until the latest release (miRBase 18), mature miRNA products arising from the "passenger" strand of the hairpin were denoted as miR* (miRNA star). Due to an increasing number of evidences showing that miR* sequences are functional, latest release of miRBase (miRBase 18) has adopted -5p/-3p nomenclature. In this thesis both miR* and -3p/5p suffixes are used to denote mature miRNA products.

1.2 miRNA detection methods

Primary discovery of miRNAs has been through traditional, low-throughput methods such as, northern blotting, cloning and PCR. Detection of miRNAs is particularly difficult and calls for a careful probe design due to their unique set of characteristics. The short length of miRNAs (22nt) make it difficult to design primers. Further, the heterogeneity of the GC-content of miRNA sequences offer them a wide range of melting temperatures. In addition, it is also difficult to distinguish between the pri-miRNAs and the precursors because of the mature miRNA sequence being present in both. Lastly, high sequence similarities of miRNAs with its family members, often to the extent of only single nucleotide, make it very hard to design detection techniques that are specific and sensitive. Available options for miRNA detection are northern blot, in situ hybridization, qPCR, microarrays and sequencing technologies. Northern blotting and in situ hybirdization (ISH) are low-throughput methdos, whereas qPCR, microarrays and NGS are the high-throughput techniques. The latter are described in Chapter 4.

1.3 Silencing through miRNAs

Mature miRNAs bound to the RISC complex act as a template for recognizing target genes. When the target gene is found, its down regulation is achieved by three alternative mechanisms: mRNA cleavage, translational repression and mRNA degradation. The choice of effective mechanism depends on the extent of complementarity between miRNA and its target site. With the perfect or near perfect complementarity between miRNA and its target, the site specific cleavage is carried out exclusively by Ago2 [74]. With an imperfect complementarity, mRNA repression or degradation takes place in the presence of four Ago proteins [180]. The mature miRNA forms a complex involving, mostly, 3' UTR region of its target and RNA-induced silencing complex (RISC). Such an association of RISC to the target mRNA results in a post-translational repression or RNA degradation.

miRNAs have been known to bind to the 3' UTR of mRNAs to direct their posttranscriptional repression. The seed region, starting from 2-8 nt in the mature miRNA, provides the specificity of target recognition. There always exists a complementarity between miRNA and its target site in the seed region. However, target recognition based only on the seed match is not sufficient for repression [71]. Length of the target site to which a miRNA binds, and the degree of complementarity between the two is also important to determine the extent of regulation. For example, repression of transcripts with 8 nt (8-mer) target site has been seen more effective compared to 7-mer sites. In addition to the degree of complementarity and length of the target site, various other features in the genomic context determine the efficacy of miRNA targeting. MiRNAs have also been shown to exhibit cooperative

regulation at their shared target site. Having multiple target sites increases the likelihood of target repression. Furthermore, the proximity of target sites on the 3' UTR, proximity of target sites of multiple miRNAs and the degree of cooperativity of miRNAs are important factors influencing the efficacy of miRNA binding. Strength and effectiveness of the binding is also determined by G:U (wobble) pairing at the 3' end of the seed region along with the imperfect complementarity between miRNA:mRNA complex [135]. Further, the evolutionary site conservation, and AU content around the target site also determine miRNA target specificity and the efficacy. Although there is an evidence for miRNA targeting the 5' UTR or the open reading frames (ORF), gene regulation by the binding of miRNA with 3' UTR is more common. Also the length of 3' UTR, as well as proximity of target sites to the ends of UTR, have positive impact on the efficacy of miRNA targeting, perhaps due to the increased accessibility of the translation machinery at the UTR ends.

1.4 MicroRNA target prediction

MicroRNA are expected to modulate the expression of hundreds of genes. Since the transcriptional repression of miRNA targets and dysregulation of miRNAs themselves have been linked to cancer and various other disorders, it is of particular interest to accuarately identify targets of miRNAs. Experimental approaches of target identification are expensive and resource intensive [200, 167, 64]. Therefore, accurate computational tools for miRNA target prediction are necessary.

Current challenges in the computational target prediction arise from the imperfect complementarity between miRNA and target within the seed region and the wobble (G:U) pair. Initial approaches for target prediction relied on a near-perfect complementarity between mRNA and miRNA in the seed region, i.e., position 2-8 in 5' end of mature miRNA. To reduce the number of false positives in the predictions, these algorithms used evolutionary sequence conservation across multiple species. Early algorithms based on seed pairing and evolutionary conservation includes, TargetScan [61], miRanda [11] and PicTar [104]. Later on, miRNA targeting events with an imperfect target-site complementarity and non-conserved target-site were reported and characterized. This lead to implementation of new algorithms that allowed for moderately stringent seed pairing and included other determinants of miRNA target recognition.

The extent of seed complementarity allowed by most of the target prediction tools ranges from perfect match (TargetScan), single mismatch in the seed region (PicTar) and multiple mismatches in the seed (miRanda miRBase [69], PITA [95]; miRWIP [76]). A weak base-pairing to the 3' end

of miRNA could compensate for an imperfect binding of 5' end of miRNA-target complex; and the target sites often reside in non-conserved regions (miRWIP [76], RNA22 [135], PITA). Recently, more careful approches to target prediction have been developed that integrate the evaluation and modeling of various sequence features surrounding the target sites, for example, miRWalk [43] and miRSVR [10]. Experimentally validated sets of miRNA targets are limited, but can be retrieved from MirWalk [43] or miRecords [192].

Unfortunately, the degree of overlap between the predicted targets from various algorithms is very low. False positive rates for some of the target prediction tools are estimated to be about 30% [25]. Further, many algorithms often have inherent biases in predictions that come from their target prediction criteria. Therefore, relying on the predictions from a single tool may result in an increased number of true negatives. For both studies described herein, I collected predictions from multiple tools (5 to 7) and accept a gene as a predicted target only if it is predicted by more than one prediction tools.

Some of the recent approaches integrate data from miRNA, gene and also protein expression for a reliable prediction of miRNA targets. The method of inferring the functional or effective miRNA-target relationships assume that the changes of miRNA expression could be reflected by the expression changes of their targets [30, 5]. Inverse-correlations between miRNA and predicted target gene expression are often used to filter out non-functional miRNA-target pairs. In Chapter 6, Paper I, a similar approach has been followed in inferring functional miRNA-target relationships.

1.5 miRNAs and cancer

miRNAs are often located within the fragile sites or at the regions that are frequently amplified or deleted in human cancers [21]. For example, miR-15 and miR-16 are located at 13q14, a frequently deleted region in chronic lymphocytic leukemia (CLL) [21]; translocation of miR-17-92 cluster containing region in T-cell acute lymphoblastic leukemia (T-ALL) and lung cancer [128, 79, 141]; and an amplification of miR-16a carrying region in glioblastoma [85]. Further, transcriptional regulation of miRNAs by transcription factors (TFs) has been observed in various cancers. Several miRNAs are regulated by transcription factors such as MYC [47, 119, 140], TP53 [35, 193] and RAS [93]. Craig et al. [38] suggested a role of MYC repressed miRNAs in the transformation of low-grade gastric lymphoma to high-grade diffuse large B-cell lymphoma (DLBCL). Taken together, an aberrent expression of miRNAs is strongly associated with development and disease transformation in cancers.

Dysregulation of miRNAs is very likely to affect expression of their target genes. Depending on the function of miRNA targets, for example in the cell growth, apoptosis and tissue differentiation, miRNAs assume oncogenic (oncomiRs) or tumor suppressive roles. For example, tumor suppressor miR-15/16 target BCL2 and induce apoptosis of cancer cells. These miRNA are often down-reguated in CLLs [21]. On the other hand, let-7 family miRNAs are tumor suppressor miRNA that modulate the expression of oncogenes such as RAS [90, 109]. Also, depending on the tissue type and type of genes they regulate, miRNAs may assume both, tumor suppressive and tumor proliferative roles [163, 80].

miRNAs have recently found a great potential in diagnostics and therapeutics [91]. Anti-miR therapies to specifically target oncomiRs or such miRNAs offering drug resistance are being developed [117, 121] and few of them are already in the clinical trial phase (Santaris Pharma, `ClinicalTrials.gov`). miRNA mimics, also known as "miRNA sponges" have been exploited to take away oncogenic effects of miRNAs (reviewed in Ebert et al.). Drug resistance in various types of cancers have been attributed to aberrant expression of miRNAs. For example, miR-221/222 expression is associated with resistance to tamoxifen in ER-positive breast cancer [132]; miR-214 and miR-125b offer cisplatin resistance by targetting PTEN and Bcl-2, respectively [194, 98].

In summary, miRNA regulation due to chromosomal aberrations or by upstream TF-regulators and miRNA-mediated regulation of specific genes and pathways, contribute to the tumorigenesis, disease progression and transformation. miRNAs also have a huge potential to be used in diagnostics, prognostics and therapeutics.

Chapter 2

Breast cancer

Breast cancer is one the most frequent types of malignancies among women from both, Western and Eastern worlds. In the past years, the incidence of breast cancer has been increasing due to improved life expectancy and lifestyle choices such as birth control pills, smoking, alcohol consumption and urbanization. In this chapter, I will briefly describe breast cancer, classes of breast cancer, estrogen receptor (ER) positive breast cancer and anti-estrogen treatment options for ER-positive breast cancer.

2.1 Breast cancer

Breast tissue is mainly made up of lobules, ducts and stroma. Each lobule contains a network of milk producing glands, each terminating into the papilla (nipple) via small ducts lined with myoepithelial cells [146]. The milk ducts and lobules lie within adipocytes and connective tissue as shown in Figure 2.1. Breast cancer is the malignancy that originates usually from the ducts and lobules. However, the cancer cells can spread into the lymph nodes, bloodstream and metastasize in other parts of the body.

Risk factors

Risk factors for breast cancer are numerous due to the complexity and heterogeneity of the disease. Mutations in BRCA1, BRCA2 and TP53 are the most important risk factor, increasing the lifetime risk of developing breast cancer at an early age [36].

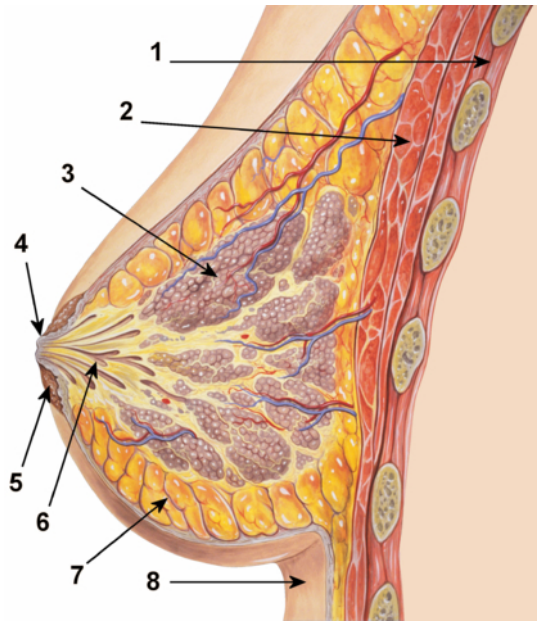

Figure 2.1. Anatomy of female breast: 1) chest wall (thorax), 2) pectorial muscles, 3) lobules, 4) nipple, 5) areola, 6) lactiferous duct, 7) fatty tissue, and 8) skin. Source: Wikipedia [124]

Other than genetic factors, lifestyle related risk factors are the following:

- Increasing age; the invasive types of breast cancers are more common in women above an average age of 55 years [56].
- Alcohol consumption, smoking and lack of physical activity [174].
- Hormone therapy after menopause.
- Obesity in post-menopausal women, due to increased estrogen synthesis from adipose tissue [156].
- Early age at menarche and late menopause.
- Oral contraceptives, as they increase the risk of breast cancer over a long term use [24].
- Mutations in tumor suppressor or oncogenes caused by an exposure to radiation or carcinogenic chemicals [174, 53].

2.2 Classification of breast cancer

Breast cancer, when put simply, is an uncontrolled growth of breast cells. However, the heterogeneity and complexity of the disease is very well known. Nevertheless, depending on the site of origin of cancer, breast cancers are divided into two main types: ductal carcinomas and lobular carcinomas. Ductal and lobular carcinomas come in invasive or noninvasive forms. The invasive ductal carcinoma (IDC) and invasive lobular carcinoma (ILC) represent the majority of breast cancers. This type of carcinoma is infiltrative, meaning that it may spread to other body parts via lymphatic system and blood stream [159]. The noninvasive types of lobular and ductal carcinomas place the patients at high risk of developing an aggressive form of the disease. Depending on the estrogen receptor (ER) status, these tumors can be treated either by anti-estrogen alone or by primary treatment combined with anti-estrogen drugs. Use of anti-estrogens in the second case is adjuvant to the primary treatment, hence the term "adjuvant therapy".

Scope of anti-estrogens

As stated earlier, the heterogeneity in breast cancer tumors makes the disease complex; therefore, the diagnosis, prognosis and treatment options are determined according to various criteria. Such distinctions are made based on the tumor stage, grade (differentiation stage), immuniohistochemical staining for receptor status as well as molecular markers obtained through microarrays or other technologies.

The microarray technology (described in chapter 4) offers a molecular sub-classification of breast cancer tumors according to their gene expression profiles. Subtypes based on the molecuar characteristics are the following: normal breast-like, basal, luminal, HER2/ERBB2 overexpressing, estrogen and/or progesterone receptor-negative-HER2/ERBB2 negative (triple negative) breast cancers.

Luminal breast cancers, also covering most of the lobular carcinoma *in situ* are more common (70% of the total cases) and are typically ER-positive [100]. Luminal A subtype, identified by high levels of ERα expression, shows a better survival [175, 176]. Whereas, luminal B subtype, characterized by low expression of ERα carries an aggressive and proliferative signature. About 50% of the HER2/ERBB2 over expressive tumors are ER-positive, but are intrinsically resistant to tamoxifen or other anti-estrogen drugs [100]. We will discuss anti-estrogen resistance in more details in section 2.3.2.

2.3 ER-positive cancer

ER is a ligand-activated transcription factor which is responsible for the development and maintenance of female reproductive organs. Apart from its primary functions, estrogen is also important for the development and maintenance of musculoskeletal, cardiovascular and nervous system in both males and females [40].

Estrogens are mainly synthesized from androgen in ovaries, but a small amount of estrogen synthesis is also present in liver, adrenal glands and the breasts. When ovaries are not functional, as in the case of post-menopausal women, estrogens are produced through a conversion of androgens by aromatase enzyme present in the organs such as adipose tissue, breast tissue, skin, brain, blood vessels and bones. Estrogens activate estrogen receptor proteins, ERα and ERβ by physical binding activities.

Activated ER binds to the estrogen responsive elements (ERE) located upstream of the ER target genes. Alternative mechanism of activated ER-binding involves various other response elements, such as AP-1 [102]. Experimentally validated transcriptional targets of the ERα include MYC, EGFR, JUN, CDKN1A, CCND1, CYP1, SPOE, FOS, TFF1, IL8 and ITGA2, etc. as listed by TRED (Ttranscriptional Regulatory Element Database) [198]. These genes are involved in the signaling events that modulate cell differentiation, proliferation and apoptosis. In the presence of acquired mutations in breast cells, an increased cellular proliferation results into breast cancer. Induced levels of oncogenes are normally the culprit for promoting proliferation of breast cancer cells.

2.3.1 Treatment of ER-positive breast cancer

Tumor size, stage and molecular characteristics, among others are the main deterministic factors for treatment options. Surgical removal of a part (lumpectomy) or entire breast tissue (mastectomy), and the surrounding lymph nodes is common for cancers in advanced stage. To reduce the risk of recurrence and eradicate any remaining cancer cells, patients are given adjuvant therapy, such as radiotherapy combined with anti-estrogen drugs.

ER-positive breast cancers are treated with anti-estrogen drugs that work by lowering the amount of estrogens in the body or by blocking estrogen from attaching to the ER. Depending on their mode of action, anti-estrogen therapy options can be classified as follows.

1. Selective estrogen-receptor response modulators (SERMs) work by blocking effects of estrogen in breast tissue by a selective binding to ERs in breast cells. Tamoxifen is the most commonly used SERM [83].

tamr.jpg tamr.jpg tamr.bb tamr.bb

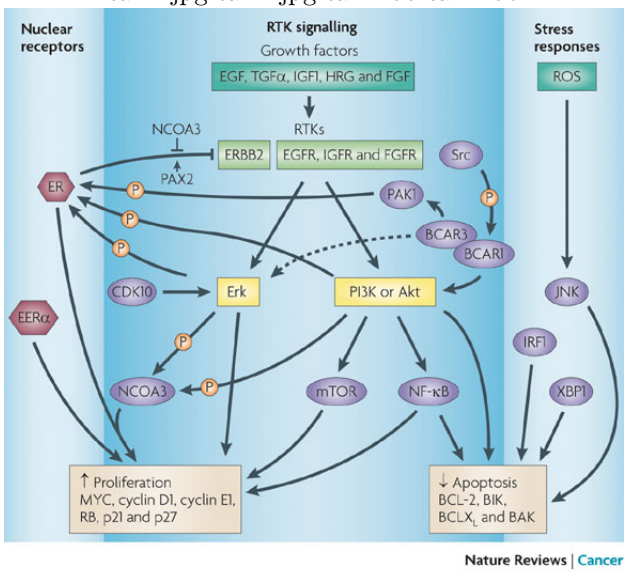

Figure 2.2. Molecular mechanisms of tamoxifen resistance and involvement of growth receptors. [139].

2. Aromatase inhibitors (AI) prevent the production of estrogen in post-menopausal women. AIs exert their function by the blockage of the aromatase enzyme responsible for converting androgen to estrogen. Because of the reduction in the availability of aromatase enzyme, less amount of estrogen is available to bind to the estrogen receptor [133].

3. Estrogen-receptor downregulators (ERDs) block the effects of estrogen by multiple modes of action. ERDs can either bind to ERs, change the shape of ERs such that they cannot bind to estrogen or reduce the concentration of ERs in breast tissue. ERDs are chosen for adjuvant therapy when the other hormonal therapies do not work. Fulvestrant (commercial name, Faslodex) is a pure anti-estrogen drug useful for the treatment of advanced ER-positive breast cancers in post-menopausal women [83].

4. Luteinizing hormone-releasing agents (LHRHs) work by stopping the production of estrogen in ovaries. As a result, a very small amount of estrogen is produced from other sources of estrogen. Such mode of action is ideal for pre-menopausal patients at early-stage of cancer [177].

2.3.2 Tamoxifen resistance

Tamoxifen, formerly known as $ICI46,474$ was the first selective ER regulator (SERM) tested in 1970. Tamoxifen blocks the action of estrogen in the breast by binding to ER, thereby making them unavailable for binding with estrogen or with co-activators of estrogen. While tamoxifen blocks estrogen in breast tissues, it mimics the action of estrogen in other tissues.

Tamoxifen has been shown to be a very effective in reducing the risk of invasive and noninvasive breast cancer by 49% compared to a group receiving placebo. The drug has also been shown to reduce the risk for women suffering from LCIS [57]. However, about half of the recurrences in ER-positive breast cancer are non-responsive to tamoxifen, either via intrinsic or acquired resistance to the drug. The primary mechanism of intrinsic or *de novo* resistance is the lack of ERα expression, however, mutations in genes such as CYP2D6 can significantly hinder the activity of tamoxifen [139, 82]. Acquired resistance to tamoxifen has been widely studied for the identification of molecular mechanims of drug resistance. Figure 2.2 shows that an aberrant activation of cellular signaling pathways involving growth factor receptors such as EGFR, ERBB2, IGFR, FGFR often activate alternative pathways to promote proliferation of breast cancer cells and inhibit apoptosis [150]. Further, changes in the miRNAs expression and miRNA-mediated regulation often lead to an aberrant activation or inhibition of signaling pathways, and thus contribute to the development of tamoxifen resistance. Roles of miRNAs in the modulation of gene expression during tamoxifen resistance are studied in detail in paper I (Chapter 6).

An analysis of gene expression profiles of tamoxifen-sensitive MCF7/S0.5 cell line and three MCF7/S0.5 derived tamoxifen resistant cell lines is described in Chapter 6. I analyzed these arrays for differential expression of genes between tamoxifen resistant and MCF7/S0.5. In Table 2.1, genes associated to various signaling pathways are listed with the cell line in which they were found to be significantly differentially expressed (abs(log_2 fold change) ≥ 0.5 and adj. p-value ≤ 0.05).

Table 2.1. Genes in the tamoxifen resistance pathway neighborhoods. Genes listed in this table were found significantly differentially expressed in the TamR-cell lines (TamR1, TamR4, TamR8) in comparison to MCF-7/S0.5 (tamoxifen sensitive) cell line [log2FC $\geq$ 0.5 and adj.p-value $\leq$ 0.05]. Theses genes showed associations with signaling pathways known to be involved in tamoxifen resistance

Cell line	Direction of change	Gene names
TamR1	Up	APOD, BMP7, C17orf91, CD59, CDKN2B, CLU, COL5A1, CPE, DUSP6, EFNB2, EPAS1, FZD7, FZD8, IRF6, ITPR1, LYN, MAGED1, MAP1B, MYO1B, NOG, NRP1, PCDH8, PLAU, PMEPA1, PRICKLE2, PRKCA, PRSS23, SDC2, SH3BGRL, TCF7L1, TGFB2, THBS1, TNIK
	Down	ARHGEF6, C9orf3, CACNA1D, CAP2, CDH18, EGR3, ELOVL2, ERBB4, ESR1, GREB1, GULP1, KIAA1324, MGP, NFKBIA, NPY1R, RBM24, SGK1, SLC3A2, SULF1, TM4SF1
TamR4	Up	ACADL, ARMCX1, ATP9A, C3, CAMK2B, CD36, CITED1, CLIC3, CPE, CPT1C, CTGF, DDX58, EGR2, EPHA4, FDFT1, FYN, GLA, GRB14, IFI27, IFIT1, IFNB1, IFNGR1, IL15, IRF7, IRF9, ISG15, JUP, OASL, PLA2G16, PLCXD3, PMEPA1, PRKD1, PTGES, RAPGEFL1, RPRM, STAT1, SULF1, TGFBR3, TLR2, TNFAIP3
	Down	ADIPOR2, AURKA, AURKB, BAMBI, BCL2L12, C12orf32, CCNA2, CCNB1, CCND1, CDCA8, CDKN1B, CENPA, CENPF, CTNNAL1, CXCL12, DLGAP5, EIF4EBP1, ELOVL2, FOXC1, GAB2, GNB4, GREB1, GULP1, HMMR, IGF1R, IL6ST, IRS2, KIF23, LIN7A, MT2A, NCAPD2, NCAPG, NCAPH, NDC80, NEK2, PCDH7, PCK2, PPM1E, PRKAR2B, RAB11FIP3, RAB31, RACGAP1, SMC2, SOX3, SVIL, TFAP2C, TLE1, TPX2, TSPAN5, WWP
TamR8	Up	CHFR, COL28A1, CP, BLNK, CALML5, CDKN2B, CROT, EXOC3, FGF13, FZD1, HIP1, IL7, IMPA1, INPP4B, LEPR, NSUN2, PCDH8, PRLR, SUGT1, TP53TG1, TPD52, YWHAG
	Down	BMPR1B, DUSP2, GNB4, ID1, IRS2, OSMR, PPM1E, SCIN, TFF1

Chapter 3

Lymphomas of the eye

Ocular lymphoma is the lymphoma localized in the eyelids, conjunctiva, lacrimal gland, intraocular or in the orbital regions of the eye. Although relatively uncommon, they account for 5-10% of the all extranoda lymphomas [60]. This chapter provides a brief introduction to the mucosa associated lymphoid tissue lymphoma (MALT) and diffuse large B-cell lymphoma (DLBCL) of eye.

3.1 The eye

An eye consists of many components, as seen in the Figure 3.1. The cornea and eyelid shield the eye from intense light. The conjunctiva is a mucous membrane covering the most of the eye. Light enters the eye through the pupil. The pupil is a hole located in the center of the iris that allows light to enter the retina. The iris controls the aperture of the pupil. The retina, located in the inner most part of the eye, contains rod and cone cells that are capable of the perception of depth and differentiation of colors [149]. When stimulated with light, the retina initiates a number of chemical and electronic modules that translate the light into signals that are processed by the brain. A uvea consists of iris, ciliary body and choroid, the vascular layer between retina and sclera. The structures adjacent to the eye, such as eyelids, eyelashes and lacrimal glands and conjunctiva comprise the ocular adnexa.

3.2 Lymphomas of eye

Hodgkin and non-Hodgkin lymphomas originate from lymphocytes, the white blood cells. In the presence of Reed-Sternberg cells in the lymphoma

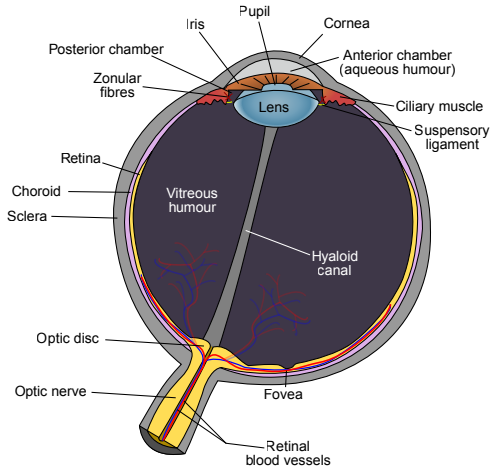

Figure 3.1. Schematic diagram of human eye. [Source:`www.wikipedia.org`]

samples, the disease is classified as Hodgkin lymphoma (HL) and otherwise it is non-Hodgkin lymphoma (NHL) [101]. The Non-Hodgkin lymphoma (NHL) is a term used for a variety of lymphomas depending on the age, grade of tumor, growth pattern and immunophenotypes [181, 88]. NHL is the most common type of ocular lymphoma. Depending on the site of origin in the eye, the ocular lymphoma can be either intraocular or orbital adnexal.

3.3 Mucosa-Associated Lymphoid Tissue Lymphoma (MALT)

The intraocular lyphomas involving retina or uvea are malignant neoplasms. Ocular adnexal lymphomas (OALs), on the other hand, describe a heterogeneous group of malignancies representing 6-8 % of the extranodal non-Hodgkin lymphomas (NHL) [189, 60]. The mucosa-associated lymphoid tissue lymphoma (MALT), which is also known as extranodal marginal zone lymphoma (EMZL) is the disease common among the elderly population. MALTs are initially indolent in nature and progress slowly, without a manifestations of symptoms for up to a year. The treatment options for OALs include radiation therapy, radiotherapy, and often a systemic chemotherapy. However, when untreated, this disease can transform into the malignant diffuse large B-cell lymphoma (DLBCL).

Pathogenesis of ocular MALT has been linked with antigenic stimulations such as from *Chlamydia psittachi*. However, there is a great amount of disagreement between studies, possibly due to the geographical differences [54, 155, 138, 39, 28, 183]. Recent studies have also implicated a number of chromosomal aberrations in MALT; for example, API2/MALT1-(t(11;18)(q21;q21), IGH/Bcl-10 - t(1;14)(p22;q32) and IGH/MALT1 - t(14;18)(q32;q21) . The t(11;18) has been found more frequently in pulmonary and gastrointestinal MALT lymphomas, whereas t(14:18) is the most frequent translocation in ocular MALTs [179].

3.4 Diffuse Large B-cell Lymphoma (DLBCL)

Swerdlow defines DLBCLs as a neoplasm consisting of a diffuse growth pattern of large neoplastic B cells with nuclear size equal or twice that of normal lymphocytes or equal to or larger than the size of normal macrophage nuclei. Like MALTs, DLBCL usually occurs at an average age of 70 years, but the disease has been seen in adults as well as children. [181]. DLBCLs involving ocular adnexal regions are rare, comprising of 13% of ocular adnexal lymphomas [55, 169]. A standard treatment option for DLBCL is CHOP, a therapy that is a combination of chemotherapy, steroids such as rituximab and a monoclonal antibody.

DLBCLs are quite heterogeneous in terms of morphology, immunophenotypic, cytogenic and genetic profiles. An immunohistochemical expression of CD10, Bcl2, Bcl6 and MUM1 is often used together with gene expression profiling to classify DLBCL into, activated B-cell-like (ABC), germinal center B-cell like(GCB) or non-GCB types, depending on the similarity of gene expression to germinal center B cells and chromosomal translocations involving BCL2/IGH locus. FOXP1 gene has been characterized as a marker of ABC, whereas amplification of miR-17-92 cluster and deletion of tumor suppressor PTEN have been linked with the GCB subtypes [110] . Classification of DLBCLs into distinct subgroups is commonly done to understand and defined underlying genetic mechanisms and to decide different therapeutic options.

Chapter 4

Measuring gene expression

Transcriptome is the term often used to describe a set of transcripts present in a given organism, tissue type, developmental stage or physiological condition. The types of RNA transcripts covered by transcriptomics include mRNAs, small RNAs such as miRNAs, piRNAs, siRNAs, snoRNAs etc., as well as long noncoding RNAs. In order to understand transcriptomics and gene regulation events involving various RNA species, it is necessary to quantify set of transcripts in the cell corresponding to a physiological condition or stage of an organism. Various technologies have been developed for the discovery and quantification of these transcripts. This chapter briefly describes some of technologies and platforms as tools of measuring gene and miRNA expression for the projects included in this thesis. I begin with an introduction of the microarray technology and describe both, gene and miRNA expression array platforms. qPCR technologies with the focus on Exiqon's miRNA qPCR are also discussed. Lastly, the small RNA-seq technology and Illumina platform are discussed .

4.1 Microarray technologies

The microarray technology works with the basic principle of hybridization between two complementary nucleic acid sequences. Fluorescence labeling of the samples is used to detect and quantify the amount bound. As opposed to Northern blots, microarrays offer a high throughput detection of transcription of thousands of genes simultaneously. Several methods based on such concept have been reported [19, 118, 184], but the most successful among them is the DNA microarray technology, implemented as oligonucleotide or spotted arrays [164].

Oligonucleotides (oligos) are the short DNA sequences (20-60nt) that are designed to match specific regions of genes. When oligos are printed or *in situ*

synthesized on solid surfaces, such as glass or silicon using photolithography and combinatorial chemistry, they are referred to as *in situ* synthesized oligonucleotide arrays. Spotted arrays, on the other hand, are prepared using a robotic array printers to "spot" small amounts of probe to desired location on the array. As the technology has evolved, a number of vendors such as Affymetrix, Agilent, Roche, Exiqon, etc. provide services for gene expression profiling of protein coding and non-coding genes. For this thesis, Exiqon's LNA microRNA microarrays and Affymetrix GeneChip Human Genome U133 Plus 2.0 arrays were used. In the following sub-sections, I will discuss both of these platforms.

Exiqon miRCURY array platform

Specific detection of microRNA expression is particularly difficult due to their short lengths, heterogeneous GC content (20-90%) and the wider range of melting temperatures (Tm) for their nucleic acid duplexes. To complicate the matters further, an extremely high sequence similarity between miRNAs of the same sequence family also makes the design of oligos difficult. Using Locked nucleic acid (LNA), a proprietary technology from Exiqon, the RNA nucleotide is modified such that the thermal stability of the oligonucleotide duplex increases significantly. This technology also provides high melting temperatures (Tm) to the LNA probe-miRNA duplexes at high sensitivity and high specificity.

Exiqon's 5th generation miRCURY LNA arrays offer capture probes for all miRNAs in human, mouse, rat and related viruses from miRBase release version 14 [68] and control probes. Each array consists of 12 sub-arrays in 4 replicates. The spot sizes are 105μm and the distance between spots is 250μm. More than 890 human, mouse and rat mature miRNAs and spike-in controls are spotted four times on the array. The arrays can be run either as dual-color using Hy3 and Hy5 labeling dyes or as single color (Hy3 only).

Typically, total RNA from sample is isolated and purified using Exiqon's miRCURY RNA isolation kit. Hy3 (Cy3) is normally used to label the samples because it is less prone to ozone quenching compared to Hy5 (Cy5) [52]. For a common reference design where the reference is made by pooling of all samples in the study or a universal reference, the sample should be labeled with Hy3 and the reference with Hy5. Next, the labeled samples are mixed and hybridized onto pre-spotted miRCURY LNA microarrays in their hybridization chamber. After following an overnight hybridization, the slides are washed and scanned using Agilent G2505B scanner. The scanned images are then exported for data analysis such as pre-processing, normalization and further analyses.

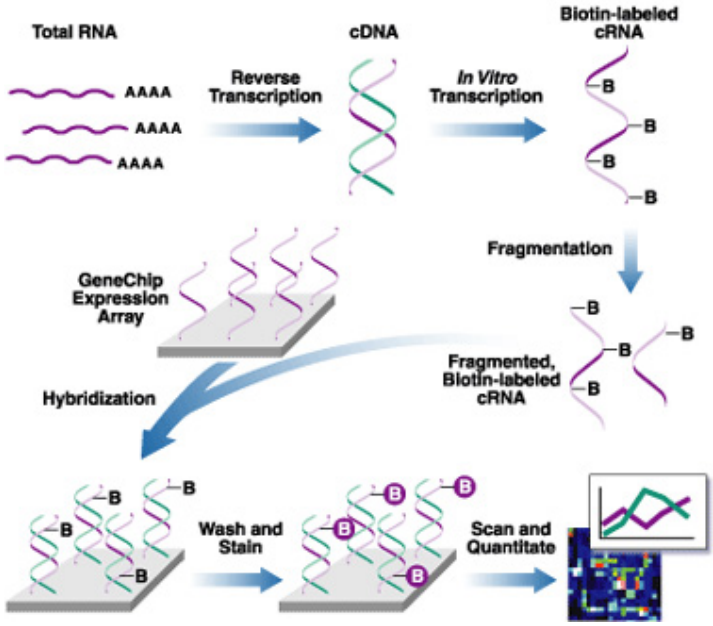

Figure 4.1. DNA microarray protocol for Affymetrix microarrays [Source : Affymetrix].

www.affymetrix.com

Affymetrix GeneChip Human Genome U133 plus 2.0

The Affymetrix GeneChip Human Genome U133 plus 2.0 array covers more than 38,000 human gene clusters (Unigene) encoding for more than 45,000 transcripts represented by about 54000 probe sets and 1.3 million 25-mer oligonucleotide probes. These probes are in situ synthesized to glass substrate by a combination of photolithography and combinatorial chemistry. The probes are derived from reference sequences from GenBank (NCBI, March 2003), dbEST (NCBI, February 2003), RefSeq (NCBI, March 2003) and NCBI human genome assembly, Build 31 [1]. Probe sequences are selected from 600 bases downstream to the 3' end of each transcript and are chosen to be unique to avoid nonspecific hybridizations as much as possible.

To prepare the sample for hybridization, mRNA is extracted, reverse transcribed to cDNA, in vitro transcribed (IVT) to cRNA, biotin labeled and fragmented. The fragmented cRNA is then left to hybridize to the array overnight (Figure 4.1). The hybridized array is then stained with a fluorescent dye, such as biotin, and on the next day it is washed and scanned via Affymetrix GeneChip Scanner 2000 [2].

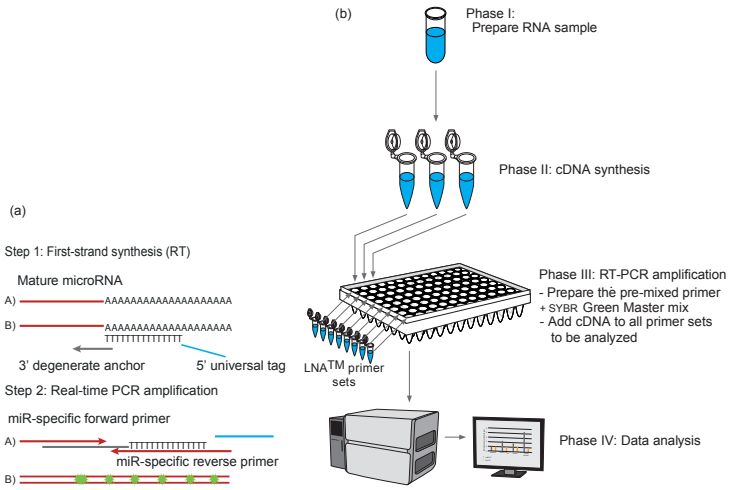

Figure 4.2. Exiqon miRNA qPCR protocol. Template RNA is first diluted in nuclease free water and later mixed with RT master mix and left for incubation to obtain cDNA. Next, the previously prepared cDNA is mixed with SYBR Green master mix, primers and added to the PCR plates. Amplification is continuously monitored for 45 PCR reactions on LightCycler480 system. [Source: Exiqon A/S].

4.2 Real-time qPCR technology

Quantitative polymerase chain reaction (qPCR) technologies offer a means to measure the abundance of genes at higher sensitivity of detection compared to other technologies such as northern blotting, RNAse protection assay or microarrays [41, 178, 66]. For microRNAs, where even a small change in expression level might be of biologically significance, a higher detection sensitivity from using qPCR technology is especially very useful.

In the real-time quantitative PCR (RT-qPCR) assay, accumulation of PCR product is measured over number of PCR cycles. The time point at which the concentration of the PCR product reaches a certain threshold level, is recorded as C_p (crossing point) or C_t (cycle time), depending on the qPCR manufacturer. In theory, the exponential amplification of template X at any given cycle n is written as:

$$X_n = X_0 * (1 + E_X)^n \tag{4.1}$$

where X_n is the number of copies of the template at cycle n, X_0 is the initial number of copies of the template and E_X is the amplification efficiency. Ideally the amplification efficiency is 100%, but in practice, it is not likely

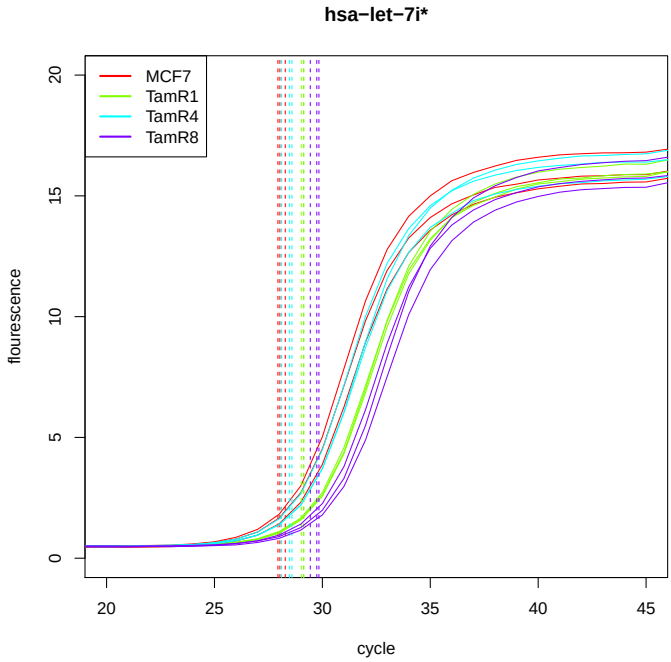

Figure 4.3. Amplification plot for hsa-let-7i* number of qPCR reactions performed on LightCycler 480. Fluorescence measured using SYBR Green I is plotted for each replicate within cell line clones, MCF7, TamR1, TamR4 and TamR8. Dotted $x-$intercepts represent the reported Cp values.

to achieve 100% amplification efficiency due to cross-contamination, thermal damage to template or also poor primer chemistry [158, 15]. Therefore, PCR efficiency is often measured by fitting a linear regression on data obtained through a dilution series.

Nevertheless, since the fluorescence is proportional to the amount of PCR product, the extent of amplification can be plotted with respect to cycle number as shown in Figure 4.3. The PCR cycle is shown on the x-axis and fluorescence from the amplification reaction on y-axis. Initially, fluorescence remains at the background level and increases exponentially depending on the amplification efficiency. In our example, C_p values are marked by intercepts on x-axis. If the template amount present at the start of the reactions is too low, more number of cycles will be required for the fluorescence to reach a specific level. Thus the reactions will have high or late C_t.

Exiqon's miRNA qPCR platform was employed for the tamoxifen resistance project .

Exiqon's miRCURY LNA Universal miRNA qPCR

Exiqon's ready-to-use Human Panel I and II, V2.R plates (product number 203608) consist of LNA miRNA PCR primer-sets that are pre-aliqoted to 384 well plates. The set of primers include more than 365 LNA primers for miRNAs, 6 primer sets for reference genes (U6snoRNA, SNORD38B, SNORD49A, miR-103, miR-191 and miR-423-p), 3 inter-plate calibrators (UniSp3 IPC) and 1 RNA spike-in control primer set. Inter-plate calibrators, pre-aliqoted on both panels, are used for calibration between PCR plates. Standard deviation between IPC replicates must not exceed 0.5. Calibration factor is defined as the difference between plate mean for IPC replicates and a global mean of IPCs from both plates. Each plate is then individually calibrated by subtracting the calibration factor from all values in the plate.

The experimental protocol consists of two steps: first-strand cDNA synthesis and real-time PCR amplification. A polyA tail is added to the mature microRNA template. The polyA tail acts as a template for binding of universal RT primer (polyT) for the reverse transcription (RT) step , The cDNA template is then amplified using microRNA specific forward and reverse LNA primers. Exiqon's miRNA qPCR assays were employed to profile miRNAs in the cell line model described in Paper I and the analysis techniques

4.3 RNA-seq technologies

Microarray and qPCR technologies, as described earlier, require knowledge of the genome *a priori*. In contrast, sequence-based approaches determine the cDNA sequence during the process, therefore not requiring the knowledge of the genome *a priori*. Sequencing approaches have existed for long time. Initial approaches were low-throughput, expensive and often not quantitative [63, 13]. Few years later, new methods such as serial analysis of gene expression (SAGE) analysis of gene expression [184] and massively parallel signature sequencing (MPSS) [147] were developed. These approaches allowed a precise quantification of gene expression. Sanger sequencing on the other hand was time-consuming and was unable to detect rare species and short species due to the limited sequencing depth [129]. Small RNAs such as miRNAs, siRNAs, piRNA, etc. are relatively hard to detect due to their low concentration in cell. Fortunately, with the advent of high-throughput sequencing technologies, it has now become possible to map and quantify small RNA transcriptomes via small RNA-seq approaches.

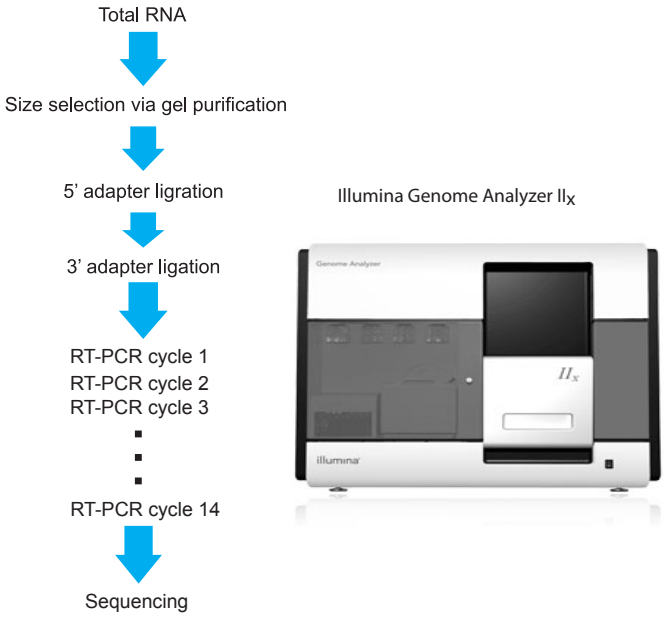

Figure 4.4. Illumina Solexa small RNA-seq protocol version 1.5.

A number of vendors, including Illumina, Applied Biosystems SOLiD and Roche 454 provide solutions for RNA-seq studies. I have analyzed single-end small RNA-seq dataset generated by employing lllumina sequencing to MCF7 and three MCF7-derived tamoxifen resistant cell line clones. In Chapter 5, analysis of small RNA-seq data and challenges in drawing statistically sound conclusions are discussed.

Illumina small RNAseq library preparation protocol

A typical small RNA library preparation begins with the isolation and purification of template RNA. As per Illumina's small RNA library preparation protocol version v1.5[86], sequencing libraries are prepared by ligating specific adapters on the 5' and 3' ends of the RNA molecules. The adapter ligated sequences are then reverse transcribed and amplified by PCR to increase the size of the library. The amplified library is then deep sequenced (Figure 4.4) on a single lane of an Illumina Genome Analyzer flow cell.

Chapter 5

Methods for the analysis of gene expression

The goal of data analysis is to highlight useful information hidden in the high-throughput gene expression datasets by inspecting, preprocessing, normalizing and modeling of data. This chapter briefly summarizes methods for the analysis of gene expression data obtained through microarrays, qPCR and sequencing technologies described in the previous chapter. Small RNA-seq and miRNA qPCR are compared in the later part of this chapter. Finally, an adaptive implementation of gene shaving, an unsupervised method of dimension reduction is described with the help of a DLBCL dataset.

5.1 Analysis of gene expression arrays

Various gene expression profiling techniques described earlier in this thesis, have made it possible to do high-throughput analysis of the abundance of transcripts. However, these experiments suffer from various technical biases that may hamper the detection of the real biological variance between the samples to be compared. Changes in the fluorescence of the dyes used, scanner settings, misaligned grids or spatial patterns introduced by the liquid diffusion patterns, etc. contribute to the technical variation. Therefore, as a first step of the data analysis, it is necessary to remove such technical biases by applying various preprocessing and normalization methods. These methods may differ based on the microarray platform used. All of the methods and analyses described here are relevant to the projects during my PhD and were carried out in R [143] environment, unless specified otherwise.

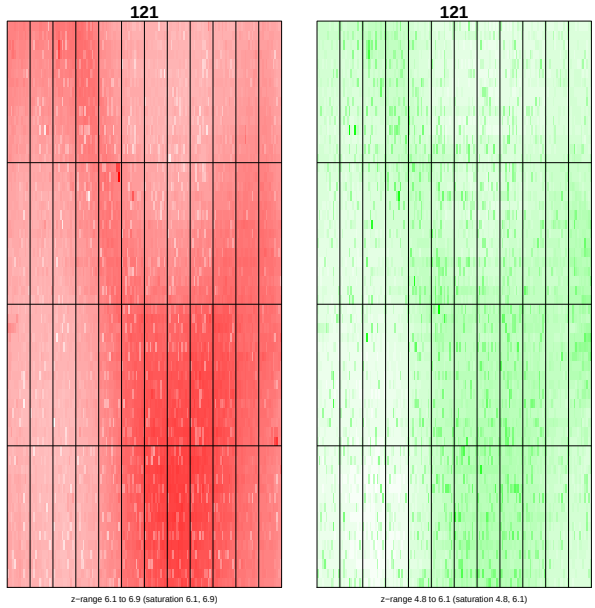

Figure 5.1. Array images from two-color array. Spatial variation of background (green channel) is evident and should be removed through background correction method.

Preprocessing and normalization methods for gene expression arrays

First step after scanning of microarrays is to assess the quality of arrays by visualization of intensities and spatial trends on each array. In case of spotted dual-channel microarrays, such as Exiqon's LNA miRNA microarrays, image analysis software records morphological details such as spot perimeter, area, location and heterogeneity measures such as standard deviations across the pixels used to construct the foreground and background intensities. Many of these tools mark poor quality spots using a practice commonly known as "flagging". However, when importing the image files into the R environment for data analysis, one can manually flag poor quality spots based on the observed differences between reported mean and median intensities for each channel separately. For example, we used a custom function that flags probes if the foreground mean and median from GenePix for a given spot differs by a threshold, 50.

```
myfun <- function(x, threshold=50) {
 okred <- abs(x[,"F635 Median"]-x[,"F635 Mean"]) < threshold
 okgreen <- abs(x[,"F532 Median"]-x[,"F532 Mean"]) < threshold
 as.numeric(okgreen & okred)
}
rawDat <- read.maimages(infiles, source="genepix.median", wt.fun=myfun)
```

Flagging does not have any effect on the intensity of spots. However, during normalization it is possible to keep these spots from affecting other good quality probes.

Background correction and normalization

In addition to the spot morphology related details, an image analysis software also returns foreground signal intensities and local background intensities for each spot on the array. In case of two-channel microarrays, the same is returned for each channel. The foreground signal measures the intensity for each spot, while background intensities relates to the noise that corresponds to nonspecific hybridizations or spatial biases on the chip. In order to obtain correct estimates of signal intensities, background correction methods are applied. (Figure 5.1). The traditional background correction method works on a simplified assumption that background intensities R_b and G_b are additive to the measured foreground intensities, R_f and G_f, respectively. Therefore, the background corrected intensities R and G can be obtained as $R_f - R_b$ and $G_f - G_b$, respectively. However, when the background intensities are higher than the foreground, corrected intensities are often negative or highly variable for low intensities spots [14].

We chose to apply normexp method for background correction in the microarray datasets used for Paper I and Paper II. The method is based on normal plus exponential convolution model as described by McGee et al [130] and used in RMA normalization method by Irizarry et al [87]. The model assumes that the observed foreground intensities result from a convolution of true signal that is exponentially distributed, and the background noise that has a normal distribution. In case of two-color array, convolution method is applied to each channel separately. $Normexp + offset$ method adds a constant offset to the corrected intensities such that the values are pushed away from zero and variance is reduced. Normexp used for my thesis has been implemented in *limma* package [170, 171] .

Next step after background correction is to apply normalization. Normalization adjusts for the technical artifacts resulting from microarray technology rather than biological differences between individual genes or samples. Normalization is applied to the background corrected log-ratios of expression, M.

$$M = \log_2(R/G) \tag{5.1}$$

The log intensity of each spot is written as,

$$A = 1/2(\log_2 R + \log_2 G). \tag{5.2}$$

One of the most common normalization strategies, central tendency (e.g. mean, median) normalization takes the global mean/median as the baseline or reference for normalization. The global mean/median is subtracted from intensities of each array. Such methods based on a global measure often

perform poorly in case of microarrays. One of the reasons for this effect is the signal dependence, requiring a signal dependent and non-linear normalization approaches. The most commonly used non-linear normalization methods are loess normalization, qspline and quantile normalization. For quantile normalization (implemented in RMA by Irizarry et al. [87]), the highest value from each array is replaced with mean of highest values, second highest value replaced by mean of second highest values, and so on. This method forces distributions of all arrays to be equal. This may not be a good normalization strategy when the arrays represent different biological conditions and assume vastly different distributions coming from biological signal. Loess and qspline normalization method apply a loess or cubic-spline fit to the average intensities on the array [190]. Qspline normalization introduces fewer errors and is more robust than loess normalization. Between-array normalization may sometimes be needed when there are substantial differences in the scale between arrays. A median absolute deviation method scales M-values of a series of arrays such that each array as same median absolute deviation.

Principal component analysis

Principal component analysis (PCA) is a frequently used procedure for dimensionality reduction of the data while retaining most of the variation in the data set. The principal components refer to the direction in which the variation in data is maximal. This method is mainly used to identify similarities between sample types or to determine grouping of data based on thee major patterns of gene expression. As outlined earlier, high-dimensional data sets can be visualized in three-dimensions by projecting samples onto the top principal components. In the R environment, PCA has been implemented as singular value decomposition (SVD) of the data matrix. In my thesis I used PCA for visualization of high dimensional gene expression data as well as for clustering of gene expression data.

5.2 Preprocessing and analysis of miRNA qPCR datasets

Melting curve characteristics

Exiqon recommends SYBR Green, a DNA-binding dye to detect the amount of amplified PCR product. SYBR Green fluoresces when it is bound to double-stranded DNA (dsDNA), including non-specific reaction products, called primer-dimers. To detect the existence of any non-specific bindings, a dissociation characteristic of the double stranded DNA is obtained as a post-PCR step.

The temperature at which a DNA strand denatures or separates when subjected to high temperatures depend upon the base composition, length

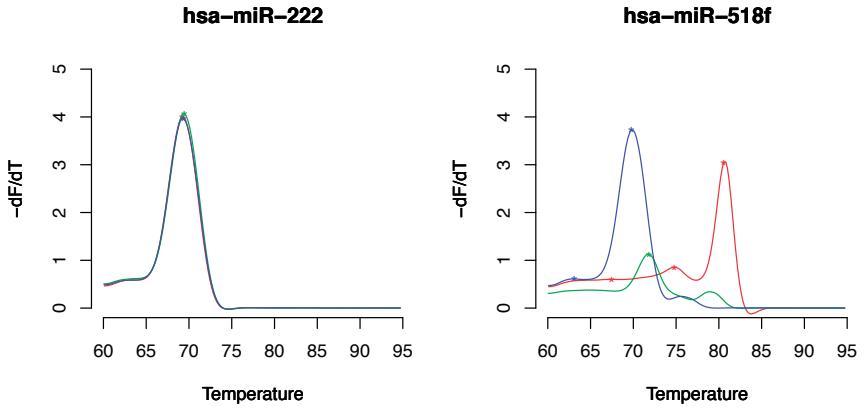

Figure 5.2. Melting curve characteristics of two miRNA PCR primers.

of the sequences and the GC content of the strand. Therefore, even a small contamination in the assay may artificially increase the fluorescence and make it impossible to exactly quantify the amount of template after amplification.

Dissociation characteristics are obtained by changing the temperature in small increments while monitoring fluorescence at each step. An increase in the temperature results in denaturation or melting of dsDNA, at which the fluorescence decreases. During cooling, the fluorescence increases as the PCR products re-anneal. The negative first derivative of the change in fluorescence over a change of temperature is plotted as shown in Figure 5.2. Melting temperatures (Tm) of the samples are displayed as peaks in the melting curves and they were identified using an R implementation of the method described by Coombes et al. [34]. This plot represents two scenarios in which the melting curves show multiple peaks or a single peak relating to non-specific or primer-specific binding of PCR products to SYBR Green respectively. Replicate assays for miR-222 show single and unique peak occurring around 70°C, whereas miR-518f replicates have multiple peaks. As a result, miR-518f assay is removed or flagged as poor quality assay. Allowing multiple peaks for an assay increases the standard error of mean, SEM, calculated over a set of replicate assays (Figure 5.3).

qPCR normalization methods

As with the microarrays, normalization of qPCR data is performed with a goal of distinguishing technical variation from true biological variation. The *gold standard* strategy of qPCR normalization relies on a set of candidate

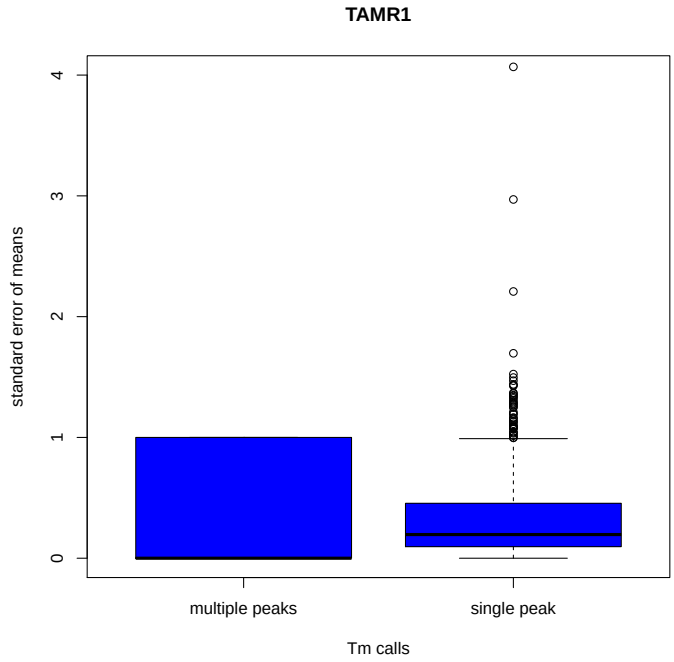

Figure 5.3. Relationship between number of peaks in melting curve characteristics and standard error of the replicate mean. A clear single peak in the melting curve characteristics decreases the probability of error.

reference genes, also called *housekeeping genes*. These genes are either pre-selected on the basis of their non-variable expression across various experimental conditions in large-scale studies, or chosen by following a bioinformatic analysis described by [182]. However, in case of high-dimensional datasets, such as miRNA qPCR dataset, use of the all data points for normalization reduces technical variation between replicates, as described by Mestdagh et al. [131].

For the miRNA qPCR dataset, consisting of four cell lines, each with three replicates (Paper I, Chapter 6), I calculated standard error of mean (SEM) for each individual miRNA across all samples within the same cell line group. As depicted by Figure 5.4, in an unnormalized dataset, raw measurements with higher C_p value have high SEM. An aim of normalization is to reduce the technical variance, modeled as SEM. I applied a series of normalizations on this dataset to assess the effectiveness of a normalization method at reducing the standard error. The following normalizations were

applied : house-keeping gene mean/median normalization, global mean/-median normalization and qspline normalization described in Workman

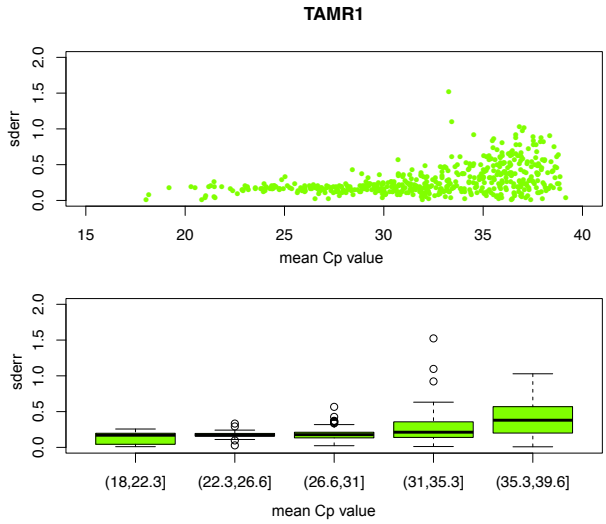

Figure 5.4. Standard error of replicate mean as function of C_p value: Before normalization. Higher variabliity among replicate measurements is noticeable at higher C_p values

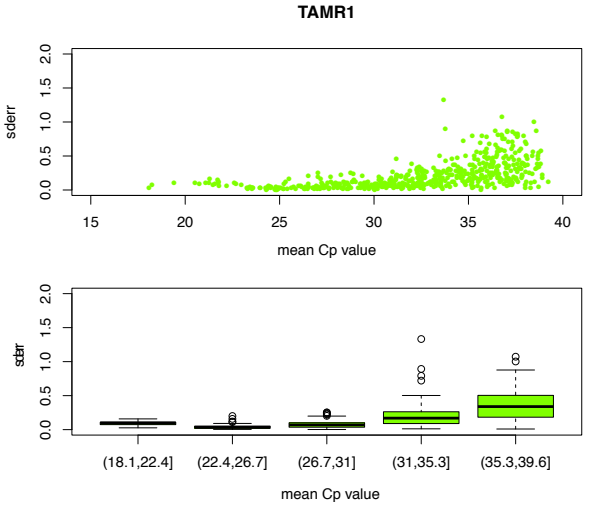

Figure 5.5. Standard error of replicate mean as function of C_p value: After normalization. Normalization is effective at reducing the standard error of means, especially for lower range of C_p values

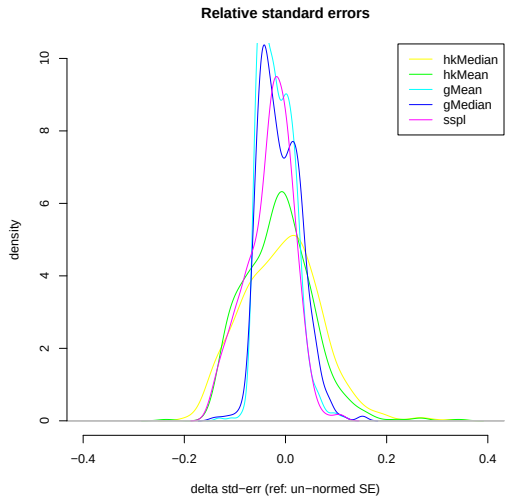

Figure 5.6. Effectiveness of normalization methods in reducing technical variation between replicates.

et al. [190]. SEMs were calculated for every miRNA across all replicates within each cell line group. Distributions of the relative SEMs are shown in Figure 5.6. As evident from the figure, both global mean and qspline method offer the highest reduction in SEMs and do not introduce additional variance as house-keeping gene based normalization methods. The effect of normalization can be seen in Figure 5.5 as reduction in variance among the replicates.

5.3 Statistical testing

Dealing with high-throughput data consisting of variety of sample types and biological or technical replicates requires appropriate statistical methods. In biology, it is often required to detect the differences in expression between two or more conditions. The first and most simple method used to compare two conditions was the use of fold-changes [29], however, this method did not take the variance within groups into the consideration and therefore, it was of limited use. Methods such as ordinary t-test [42] assume the null hypothesis that the difference in the means of two groups to be compared is zero. t-test is performed on a gene-by-gene basis. When testing multiple genes using t-test, the probability of error increases.

Many experiments often need to compare more than two groups at a time. In these cases ANOVA can be used to identify differentially expressed

genes between more than two conditions. In contrast with t-test, ANOVA estimates the variability within group based on information borrowed from all the groups in the comparison, giving more statistical power and sensitivity compared to the t-test. However, both t-test and ANOVA are parametric and make an assumption of normality and independence of variables. Wilcoxon rank-sum test (used in Paper I), on the other hand, is one of the non-parametric tests that makes no assumptions about distribution of data [188].

A method by Smyth et al [170], implemented as a Bioconductor package *limma*, applies a linear model to each gene. The model is designed according to the supplied experimental design and sets of independent variables or conditions to be compared. The coefficients of the fitted models describe the differences between groups. This method is extremely useful as it is applicable to microarrays as well as qPCR datasets. With a well designed model, the method handles technical probe replicates as well as biological replicates in a block design. For this thesis, the *limma* approach has been used to compare different experimental conditions.

Number of simultaneous hypothesis testing calls for a multiple testing correction. Several methods for multiple testing correction have been developed, in which Bonferroni and Benjamini Hochberg [7] methods are the most conservative. I have used Benjamini Hochberg method of multiple testing correction throughout this thesis.

5.4 Preprocessing and analysis of small RNA-sequencing data

Read quality control

Raw reads are received in the files formatted using FASTQ format as shown in 5.7. FASTQ format is widely used to standard format to store the read sequences and Phred quality scores in the same file [33, 50]. Although Illumina sequencing uses different quality scoring, the file format looks similar

```
@QQRBfh*^Etsgf8ZG&O$BqVMdfs
TGAGGTAGTAGGTTGTATAGTTATCTCGTNTGCCN
+HDG/SgX&peLbbqDM)R1v>J%pO6
bbbbb]bbab^J^Y]b_bbbbbbbb_bb_BBBBBB
@&-<lFHBuQZZg3p8IKOpbLvfHw6
CACCATGATGGAACTGAGGATCTGAGGATNTCGTN
+Nu%MyYz^TlvOM4=Sb3SC0q&DGw
a^Za[[^^Z^a_[^a^Z^^U^``aBBBBBBBBBBB
@t(hf23aAWaiE*7(ZRpVNu>7vjV
```

Figure 5.7. Example of fastq file

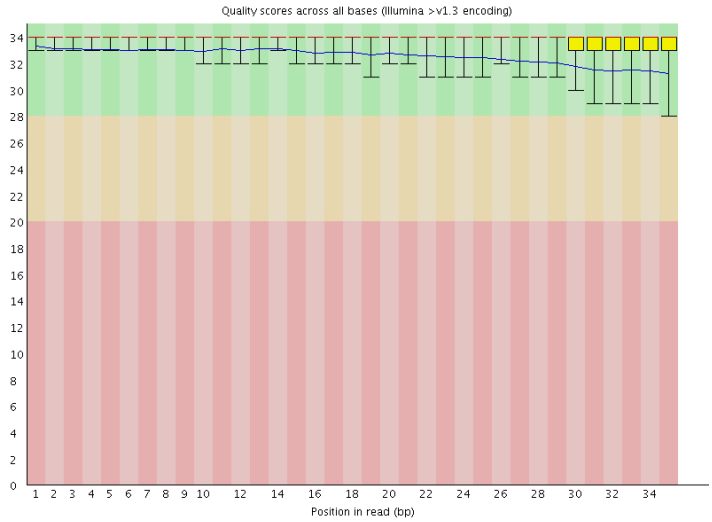

Figure 5.8. Per base quality report generated for raw fastq file using FastQC.

to FASTQ. A typical FASTQ block allocates four lines per sequence. Line 1 begins with @, and is followed by a sequence identifier. Line 2 contains raw sequence letter. This line is followed by a + sign, and an optional sequence identifier (encoded in Illumina FASTQ files). Line 4 contains quality scores in encoded in ASCII.

Phred quality score is a non-negative integer that represents error probability of a base call to be wrong. Therefore, a Phred quality score Q is written as :

$$Q = -10 * log(err)/log(10) \tag{5.3}$$

where err is the error probability of a base. Thus, to allow error probabilities up to 0.01, the minimum Phred quality score required is 20. Quality scoring in Illumina/Solexa raw files is slightly different. Illumina/Solexa quality score sQ is given by

$$sQ = -10 * log(err/(1 - err))/log(10). \tag{5.4}$$

Since various alignment algorithms assume Phred quality score to assess quality of base call, we converted Illumina/Solexa scores into Phred scores using a Python script.

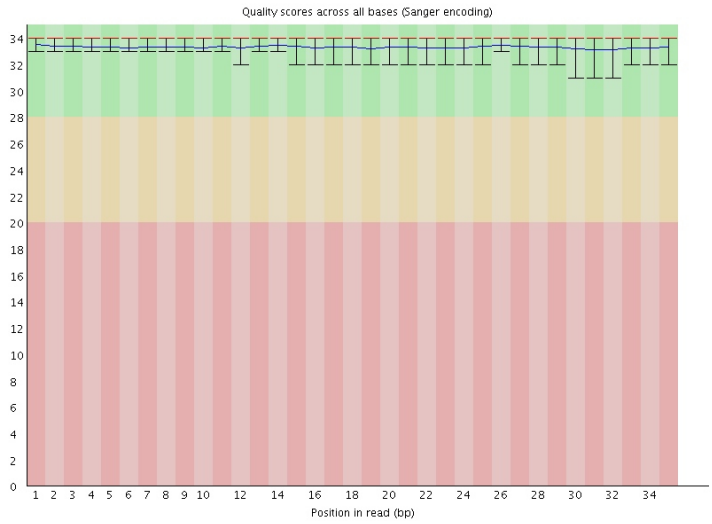

Figure 5.9. Effect of trimming low quality bases. Per base quality report generated for fastq file using FastQC.

Using a quality control tool, FastQC [165], I generated quality control reports on the reads. Figure 5.8 depicts a distribution of base quality scores over one FASTQ file. The quality of base typically declines toward 3' ends of the read. Base errors may cause mapping errors since many aligners rely on read-quality score for generating best quality mappings. Therefore, in order to achieve reliable alignments and to reduce the probability of false positives, I trimmed the 3' ends of reads and quality scores. Distribution of quality base scores after the trimming of low-quality bases can be seen in Figure 5.9.

Other parameters such as per base sequence quality, sequence length distribution and sequence over-representation were also assessed using FASTQ. Trimmed reads of lengths shorter than 15 bases were removed. Over-represented reads were assess for a match with the adapter sequences used during the library preparation. Reads that map adapter sequences were removed. After the preprocessing steps, the number of *clean* reads corresponds to about 99% of the total reads. The set of clean reads was heavily duplicated, as elucidated in Table 5.1.

Read alignment

Alignment is the process by which sequencing reads are aligned to a reference genome or transcriptome. We were particularly interested in differential expression of miRNAs between tamoxifen-resistant and tamoxifen-sensitive cell

Table 5.1. Summary statistics on sRNA-seq reads

Cell line	RNA μgrams	Total Reads	Clean Reads	Unique Clean Reads	% of read-duplication in unique clean reads
MCF7	36.912	11086223	11045089	345512	3.13
TAMR1	36.192	11334364	11296849	311877	2.76
TAMR4	35.088	11080176	10954570	467677	4.2
TAMR8	29.616	12207743	12169298	447376	3.68

lines. Various tools for aligning RNA-seq data are available, such as BWA (Burrows Wheeler transformation-based tool by [114, 115], Bowtie [105], Novoalign (`www.novocraft.com`), MAQ [116], to name the few. However, alignment of small RNAseq has its own challenges. Reads from small RNA-seq are shorter and miRNA-species to which the reads are mapped, are known for nearly high-sequence similarities in the seed-regions. Further, RNA editing [12, 191] in miRNAs can also cause some reads not to map the genome. Allowing mismatches also increases the chance of multi-mapped reads. Further, the problem of handling multi-mapped reads require a careful consideration. Some researches choose to exclude the multi-mapped reads from the analysis [157] to avoid running into a further problem of allocating those reads between possible source loci. On the other hand, when multi-mapped reads are not excluded from the analysis, allocation of these reads to multiple loci could be weighted in proportion of the number of uniquely mapped reads at each loci [113, 137].

We applied a different approach of handling read alignment as described below:

1. Trim low quality (Phred score < 20) bases at the 3' read-ends of reads based on fastq files.
2. Remove reads of length shorter than 15bp.
3. Generate a dictionary containing distinct read sequences (*unireads*). Record the copy number, cn of each unique read sequence. Let us call the *unireads* u, where $u \in U$.
4. Discard reads with $cn = 1$, as they might have resulted from sequencing errors.
5. Generate fasta files from *unireads*.
6. Prepare reference database to map the reads against Human precursor miRNA sequences obtained from miRBase [68] release September 2010.

7. Alignment of *unireads* to reference database. No gaps or mismatches were allowed. Minimum length of alignment = 15.

8. Discard *unireads* that map to multiple miRNAs (multi-mapped reads).

9. Count number of *unireads* M, mapped to each miRNA, i.

10. Quantify expression of each miRNA in terms of read counts. Expression e_i for each miRNA is obtained as follow:

$$e_i = \sum_{i=1}^{M} u_i * cn_i \tag{5.5}$$

After experimenting with several aligners, I chose to employ BLAT standalone version 34 [94]. The readcounts per miRNA were obtained as outlined earlier.

Table 5.2. Summary of RNA-seq alignment

Cell line	**Input *unireads* (cn>1) count**	***uniread* (cn>1) Aligned**	**Percentage aligned**
MCF7	83171	17695	84.87
TAMR1	78068	16163	84.21
TAMR4	118133	16957	81.47
TAMR8	115957	17067	81.20

Normalization in small RNA-seq datasets

Current status of RNA-seq normalization methods is relatively less mature compared to microarray technologies. However, it often involves adaptations from normalization strategies from microarray field. For example, the concept of housekeeping genes in microarrays translates to rank-invariant sRNA sequences that can help in establishing a baseline for normalization methods [18]. Spike-in control RNA is often added to the RNA library during the RNA preparation process, as followed by Fahlgren et al. [51]. Reads that were observed in a single copy are often removed from the dataset to eliminate outliers caused by sequencing errors. Most commonly used normalization method in small RNA-seq is to apply linear scaling by using "library size" as the normalization baseline. Tools such as DESeq and EdgeR apply sophisticated methods to measure differential expression between sample groups[153, 4]. In the absence of replicates for MCF7/TamR cell lines, log_2 foldchanges calculated by DESeq were close to the log-ratios of read counts in comparisons.

5.5 Small RNA-seq versus miRNA qPCR

Each technology has its pros and cons. Small RNA-seq provides an unbiased view of the small RNAs in a biological sample. Unlike microarrays or qPCR, sequencing does not require the knowledge of the genome *a priori*, and therefore the discovery of novel noncoding RNA species is possible. Sequencing also offers a wider dynamic range compared to qPCR. Due to all the promises made by sequencing, miRNAs in breast cancer cell lines (MCF7/S0.5, TamR1, TamR4 and TamR8) were profiled using Illumina small RNAseq, however without replication. We had an opportunity to compare miRNA qPCR results with those from small RNA-seq experiment. Before going into the details, It is important to note that the miRBase version to which reads were aligned to was identical to miRBase version on which qPCR primers were designed. An intersection of miRNAs detected by both technologies were investigated for their agreement. Raw read counts obtained from the procedure described in earlier in Section 5.4 were compared with the C_p values in corresponding cell line groups. Individual comparisons with their correlations (calculated as Pearson's correlation coefficient, PCC) in each cell line are shown in Figure 5.5. An overall correlation of -0.71 across all samples was observed. As evident from the figure, reads with low read counts corresponds to high C_p values, and miRNAs with high abundance generally corresponds with low C_p values, with few outliers.

Overlap of differentially expressed genes from qPCR and sequencing showed poor agreement. As shown by Figure 5.11, agreement of fold-changes between corresponding cell line-based contrasts (*i.e*, TamRx vs. MCF7 suggests poorer correlation PCC values 0.46 for TAMR1, 0.50 for TAMR4 and 0.56 for TAMR8.

In absence of replicates for sequencing dataset, it was difficult to assess the significance of log-ratios of differential expression, therefore a number of miRNAs detected by sequencing were not detectable using qPCR. For example, miR-452 with the log-ratio of -2 in TamR1 vs MCF7 contrast, was not measurable using qPCR. A similar trend was seen in other cell line contrasts in RNA-seq data : miR-1293, miR-206, miR-1180, miR-449c. These miRNAs, although showing differential expression in RNA-seq were not at all detectable by qPCR. Once again, in the absence of replication in RNA-seq experiments, it is hard to state whether or not the miRNAs with differential expression in RNA-seq were false-positives.

In summary, I would like to state that despite of promises of high specificity and sensitivity, and obvious pros of using sequencing technology to profile miRNAs, it is extremely important to make careful choices for the experiment design, number replicates, library preparation protocol, preprocessing and normalization method for data analyses steps.

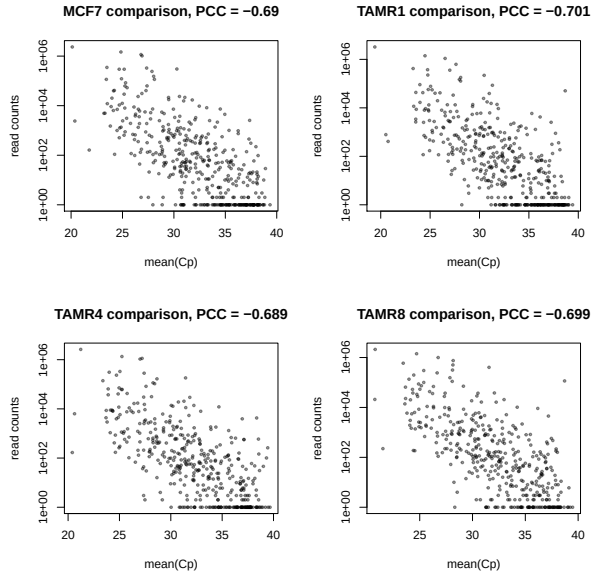

Figure 5.10. Correlations between miRNA expression levels from qPCR and small RNA-seq. $x-$ axis refers to mean C_p values within a cell line and $y-axix$ refers to raw read counts for the corresponding cell line. Values of Pearson correlation coefficients are indicated in the title.

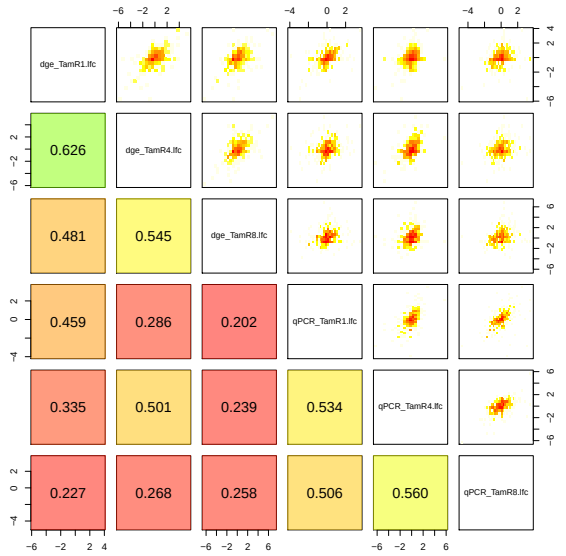

Figure 5.11. Correlation of miRNA fold changes measured via Exiqon's qPCR and Illumina's small RNA-seq technologies. Fold changes measured through RNAseq are denoted using prefix "dge", whereas those measured through miRNA qPCR are prefixed by "qPCR". Blocks above the main diagonal shows distributions of fold changes as 2D histograms, whereas blocks below the main diagonal represent Pearson Correlation Coefficients for corresponding comparison

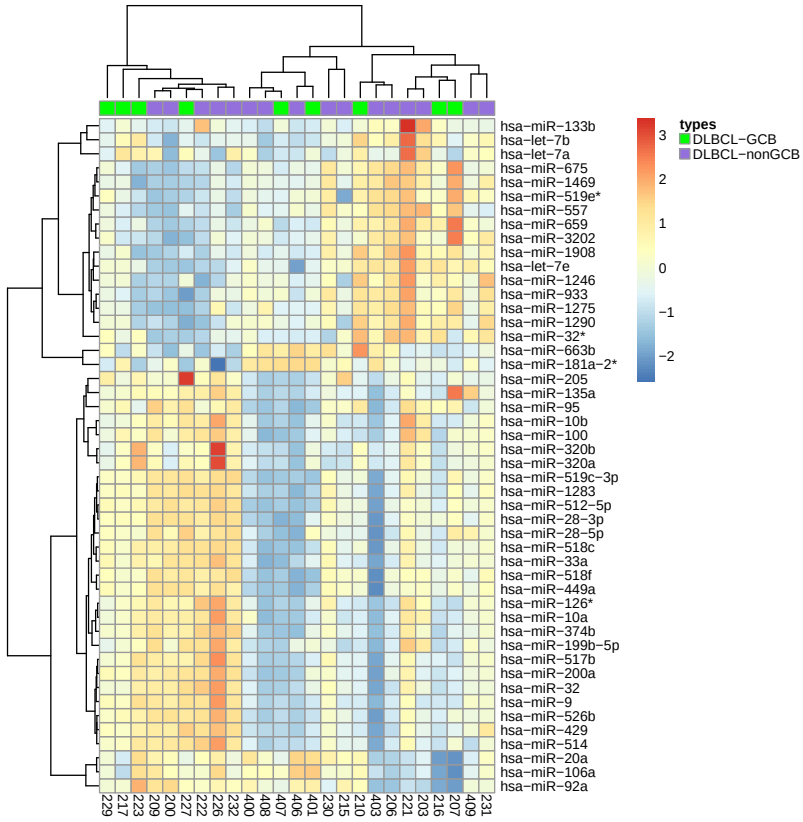

Figure 5.12. miRNA and sample subgroups identified by applying gene shaving to miRNA-expression data

5.6 Searching for patterns within expression data

Clustering technique is commonly applied to gene expression array data for grouping sets of genes or samples together according to their expression patterns. Although the classical methods, such as hierarchical clustering, provide a unique reordering of genes that define subgroups of genes or samples with unique expression patterns, but it loses the finer structures within the data. In 2002, Hastie and Tibshirani [78] developed "gene shaving", a two-way clustering method that can find finer structure in data and possibly overlapping clusters using supervised or unsupervised approach. The method works as follows. Initially, each row of a gene expression matrix, X, is centered to have zero mean. Singular value decomposition is applied and inner product of each row in X and first principal component is calculated.

Fraction number of genes, called "shaving factor" (typically 7-10%) having a lowest inner product to the leading principal component are "shaved" from the data. This process is repeated until a single gene remains. One of the gene-block of size k is selected from the nested sequence of gene clusters by using an optimizing criteria, usually a gap statistic.

I acknowledge the possibility that the first principal component may not always explain all of independent variables, and therefore removing genes on the basis of their covariance to only the first component might be risky. Therefore, we slightly modified the original method in following manner.

- Instead of computing inner products to only first principal component, we calculate inner products to first three components,
- The termination criteria for gene shaving is adaptive : iterations end when a geneset showing a threshold amount of variance remain. In other words, the algorithm terminates when it starts removing more informative genes that have higher inner product with the leading principal components, and
- We reduce the proportion of genes removed at each iteration; value for the shaving factor is reduced after a specific number of iterations has passed without termination (20 iterations).

Significance of the clustering result is tested with 100 random permutation tests by changing the row labels of X.

I was particularly interested in finding miRNA signatures that might help in defining groups of samples that may correspond to specific set of clinical characteristics. Therefore, this method was applied to a miRNA expression dataset consisting of 25 ocular DLBCL patient samples described in Paper II. Figure 5.12 shows a reduced set of miRNAs obtained through the application of an adaptive gene shaving procedure on this dataset. As evident from the heatmap, finer structures within this reduced dataset is characterized by high variation across samples and high correlation across genes. Sample subgroups were obtained on the bases of clusters of miRNAs showing similar signatures 5.13 and they were used to characterize groups of patient samples. Survival analysis on the sample subgroups defined by miRNA signatures did not provide any significant correlation to survival. This might be due to the limited number of samples (n=25) and a rather short clinical history of the elderly patients. No significant enrichment of the clinical characteristics within the subgroups was found.

In summary, although very successful in identifying finer structures within the data, application of this method to a small patient dataset with heterogeneous clinico-pathological and molecular characteristics did not yield

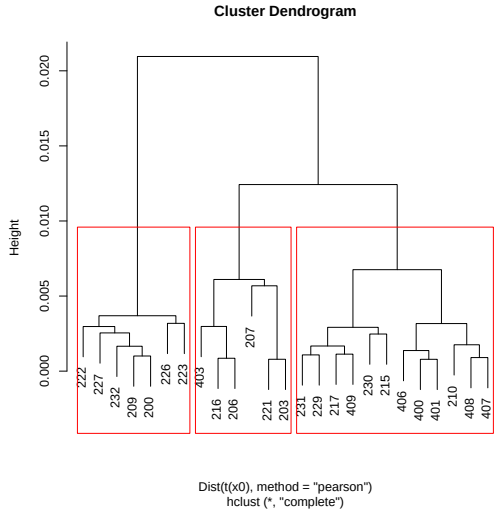

Figure 5.13. Sample subgroups identified based on miRNA signatures by following adaptive gene shaving

statistically significant associations of identified sample subgroups to survival characteristics. Further testing of the method on a large dataset consisting of heterogeneous samples may yield better and statistically significant associations to the survival or other clinical characteristics.

Part II

Papers

Chapter 6

MicroRNAs in tamoxifen resistance

6.1 Prelude

The aim of the project presented as a manuscript *"Integrative analysis of miRNA and gene expression reveals regulatory networks in tamoxifen resistance"* was to perform an exploratory analysis of miRNA transcriptome in a cell line model composed of tamoxifen sensitive MCF7 and three MCF7-derived, tamoxifen resistant cell lines. Several studies in the area of tamoxifen resistance have focused on a limited number of miRNAs [199, 132, 121]. This study, therefore was planned to gain global insights into the tamoxifen resistance conferred by miRNAs. However, to better understand the consequences of miRNA expression, a gene expression dataset, generated using the same RNA, was obtained and investigated in relation to miRNA expression profiles.

This project was a collaboration with the Sino Danish breast cancer research center, University of Copenhagen, Denmark. miRNA qPCR profiling was performed at Copenhagen University. Gene expression dataset was obtained from Daniel Elias from his unpublished work. I was solely involved in the bioinformatics analysis of data.

6.2 Manuscript

Integrative analysis of miRNA and gene expression reveals regulatory networks in tamoxifen resistance

Tejal Joshi[1], Daniel Elias[2], Jan Stenvang[3], Maria Lyng[4], Anne Lykkesfeldt[5], Nils Brünner[3], Ramneek Gupta[1*], Christopher T Workman[1*], Henrik J. Ditzel[2*]

1 Center for Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark, Denmark.
2 Department of Oncology, Odense University Hospital, Odense, Denmark
3 Department of Veterinary Disease Biology, University of Copenhagen, Copenhagen, Denmark
4 Department of Cancer and Inflammation, Institute of Molecular Medicine, University of Southern Denmark, Odense, Denmark
5 Center for Cancer Research Society, Cancer Society, Copenhagen, Denmark.
∗ Joint corresponding authors

Keywords: ER-positive, Tamoxifen resistance, miRNAs, breast cancer

Abstract

Tamoxifen is a commonly used anti-estrogen treatment for patients with ER-positive breast cancer. Despite its effectiveness, more than 30% of the patients receiving this treatment acquire resistance to the drug. Unfortunately, the underlying molecular mechanisms of drug-resistance are still unknown.Here we study tamoxifen resistance through a systematic correlation analysis of the miRNA and mRNA expression profiles and identify miRNA-mediated regulation in a cell line model composed of an ER-sensitive parental MCF-7/S0.5 and three derived tamoxifen resistant breast cancer cell lines. We observed a significant differential expression of a number of miRNAs in drug resistant cell lines, suggesting their role in the development of drug resistance. Genes regulated by these differentially regulated miRNAs suggest that multiple mechanisms of miRNA-mediated tamoxifen resistance exist, encompassing various receptors signaling pathways, aurora B kinase signaling and FOXM1 transcription factor (TF) activities. Further, our analysis identifies evidences for miRNA-mediated regulation on ESR1,

PGR1, FOXM1 and 14-3-3 family genes. With the integration of such inferred miRNA-target relationships and existing knowledge on TF-mediated regulation of miRNAs and protein coding genes, we provide hints toward FOXM1-mediated regulation of miR-135a, in the view of tamoxifen resistance.

Introduction

Tamoxifen is widely used as adjuvant therapy for ER-positive breast cancer in both pre- and post menopausal women. Tamoxifen has been shown highly effective in 26% reduction in mortality after 10 years of treatment [45] as well as in reducing the incidence of breast cancer in patients with high risk of developing the disease. However, almost all patients in metastatic disease and more than 35% of the patients receiving tamoxifen as an adjuvant therapy undergo relapse [151].

Ideally, during the initial years of the treatment, tamoxifen acts by disrupting the activity of estrogen by forming a bond with ER-alpha, thereby preventing the binding of ER-alpha to estrogen receptor. Previous studies have shown a loss of ER-alpha expression in ER-positive breast cancer patients treated with tamoxifen [151],[122]. While the cause of such altered expression of ER-alpha might be linked to epigenetic changes such as, hypermethylation of CpG islands in 5‘-regulatory regions of ER-alpha, other genetic events that affect expression of ER-alpha also exist [65]. Various attempts have been made to characterize tamoxifen resistance via altered expression of ER-beta, co-activators and co-repressors of ER-alpha, IGF1, IGF1R, IGFBP3 and various other genes [151, 139, 90, 127, 126]. However, events leading to such regulation of gene expression have not yet been completely understood.

MicroRNAs (miRNAs) are small, approximately 21 nucleotide long, non-coding genes that regulate gene expression at the post-transcriptional level. In mammals, more than 60% of protein-coding genes undergo post-transcriptional regulation in the form of degradation or translation inhibition through miRNAs. Due to their potential in regulation of tumor suppressor or tumor promoting genes, miRNAs have recently garnered attention. The let-7 family of miRNAs have tumor suppressive effects because they target oncogenes via post-transcriptional regulation; whereas miR-155, miR-17-5p, miR-21 are known for their oncogenic roles in tumors [97, 145, 70]. In cancer cells, drug-induced aberrant expression of miRNA may change drug sensitivity by modulating gene expression of their target genes. Gene regulation by miRNAs often results in activation or dysregulation of various pathways that might be crucial for development of drug resistance or restoration of drug sensitivity. MiR-214, miR-200c and miR-376c are critical factors for drug resistance in a variety of cancers [23, 194, 196, 75]. For example, miR-221/222 regulates levels of ER-alpha, and thereby plays a critical

role in ER-positive cancers and tamoxifen resistance [199, 132]. A recent study also showed that miR-22 offers regulation of ER-alpha expression to an even greater extent[142]. Suppression of miR-222 and miR-181b has been shown to restore drug sensitivity in tamoxifen resistant xenografts in mouse due their targeting effect on TIMP3 [121]. Most of these studies addressed only a limited number of miRNAs and hence they do not provide a global view of miRNA-mediated drug resistance. Therefore, a systematic study of miRNA-target relationship on a cell line model may provide us explanations as to why certain patient groups acquired resistance to the drug.

In present study investigate miRNA-mediated gene regulation using a tamoxifen sensitive MCF-7/S0.5 cell line and three MCF-7/S0.5-derived tamoxifen resistant cell lines, TAM^R-1 (TamR1), TAM^R-4 (TamR4) and TAM-^{R}-8 (TamR8). Since it is known that miRNA target predictions often suffer from high number of false-positives, we performed an integrative inverse-correlation analysis of miRNAs and gene expression profiles to distinguish genuine or functional miRNA-target relationships from miRNA target predictions. The results provide interesting insights into affected biological processes during the drug resistance. Further, using a systems biology approach of integrating various levels of gene regulation, we provide mechanistic hypotheses regarding transcription factor (TF) mediated miRNA regulation and its impact on drug resistance.

Materials and Methods

Cell lines

The human breast cancer cell line MCF-7 was originally received from The Breast Cancer Task Force Cell Culture Bank, Mason Research Institute (Worcester, MA). The MCF-7 cells were gradually adapted to grow in low serum concentration and the tamoxifen sensitive subline MCF-7/S0.5 [17] was used for establishment of tamoxifen resistant cell lines by long term treatment with 10-6 M tamoxifen as described in [123]. Three tamoxifen resistant cell lines TAM^R-1 (TamR1), TAM^R-4 (TamR4) and TAM-^{R}-8 (TamR8) were derived from distinct colonies emerged in a culture with tamoxifen treated MCF-7/S0.5 cells [123]. The cell lines were grown in standard phenol-red-free DMEM:F12 (1:1) medium (Gibco, 21041-025) supplied with 1% heat-inactivated FBS (Gibco, 10270-106), 6ng/ml insulin (Sigma, I6634) and 2.5mM glutamax (Gibco, 35050). The standard medium for the tamoxifen resistant cell lines was supplied with 10-6M tamoxifen (Sigma, T5648). MCF-7/S0.5 and the tamoxifen resistant cell lines were passaged once every week and seeded out with 1x105 and 1.4x105 / T25 flask, respectively. The cell lines were cultured at 37°C and 5% CO2 and kept at low passage numbers throughout the experiments (<10 passages) to avoid phenotypic changes.

miRNA quantitative real time PCR

Total RNA from cell line cultures was extracted and purified using TRIzol (Invitrogen supplied) and EtOH precipitated. The microRNA qPCR profiling was conducted by the miRCURY LNA® Universal RT microRNA PCR system (Exiqon, Denmark) according to the manufacturer's instructions. In brief, 50 ng total RNA was reversed transcribed in 40 μl reactions on a Bio-RAD S1000 Thermal Cycler (60 min at 42°C followed by heat-inactivation of the reverse transcriptase for 5 min at 95°C). From the resulting cDNA 32.5 μl was used for a SYBR green master mix and run in 10 μl real-time amplification on microRNA Ready-to-use PCR, Human panel I+II, V2.R plates (Product number 203608) on a Roche LightCycler® 480 real-time PCR system. The applied PCR settings were: 10 min at 95°C, 40 amplification cycles (95°C/10 sec, 60°C/1 min, ramp rate 1.6°C/sec). For each cell line three independent samples were analyzed.

Gene expression microarrays

Gene expression arrays were obtained from Elias et al [48]. The authors profile gene expression of MCF-7/S0.5, TamR1, TamR4 and TamR8 using GeneChip® Human Genome U133 plus 2 Arrays (Affymetrix) in 6, 3, 3 and 2 replicates, respectively. These arrays were profiled by following the RNA preparations identical to that of the miRNA qPCR assays.

Analysis of miRNA qPCR data

Primers with more than one peak in its melting characteristics were identified and annotated via melting curve analysis performed in R.
Expression levels (Cp values) of the housekeeping genes (miRNAs or snoRNAs) on the qPCR assays expressed a greater degree of variance (equivalent to more than one Cp value) across all samples. Further, Cp values of housekeeping genes did not span the entire range of Cp values observed in the experiment. Therefore, we applied a signal-dependent non-linear normalization method, similar to the one described in Workman et al in 2002 [190]. The offset amounts were calculated by fitting cubic-spline to the entire dataset and thereafter added to the raw data to achieve desired target distributions. Our normalization approach offered 20% reduction in replicate standard-error of means. As a next step, we used limma [170, 171] from Bioconductor package to determine differential expression of genes for following contrasts: TamR1 vs. MCF7, TamR4 vs. MCF7 and TamR8 vs. MCF7; henceforth referred to as TAMR1, TAMR4 and TAMR8, respectively.

Analysis of microarray data

Raw data was obtained from Elias et al [48]. Our inverse correlation analysis to infer functional miRNA-target relationships is highly dependent

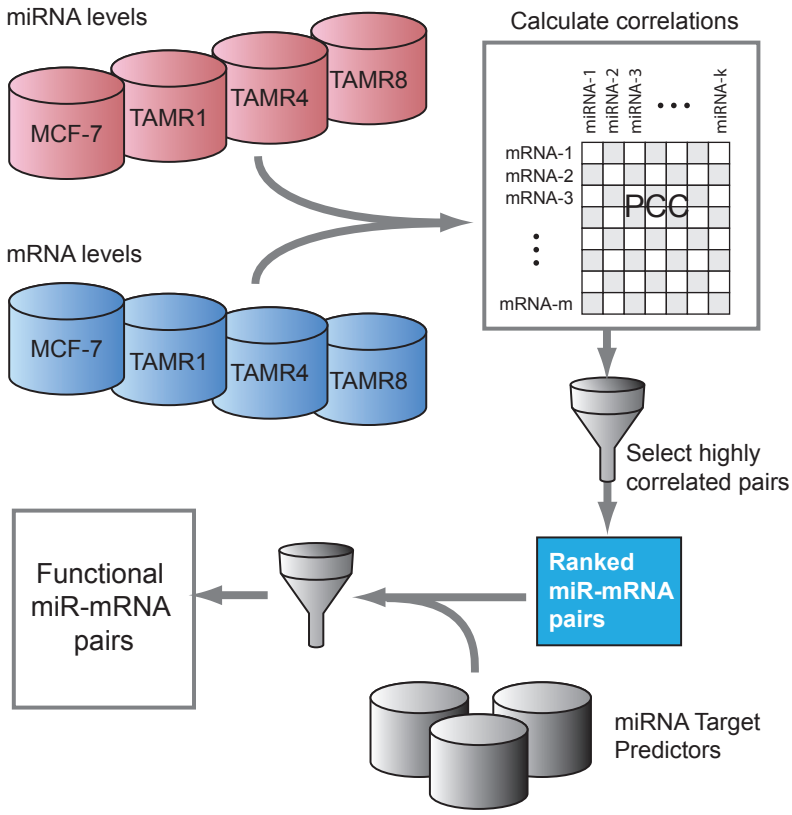

Figure 6.1. Schematic workflow of the inverse-correlation analysis on miRNA-mRNA expression to identify functional miRNA-targets

on the gene expression patterns. Therefore, to reduce technical variance as much as possible, raw intensities were reanalyzed using RMA (robust multi-array average) [87] implemented in Bioconductor package affy [62]. RMA normalization consists of three steps: background correction, quantile-normalization at probe-level and probeset summarization using a robust linear model fit to the log-transformed normalized values. Differential expression analysis was performed using limma [170, 171] package in R for the following contrasts: TamR1 vs MCF7, TamR4 vs MCF7 and TamR8 vs MCF7; henceforth referred to as TAMR1, TAMR4 and TAMR8, respectively.

Integration of miRNA-mRNA expression data

An unbiased set of miRNA targets was compiled by taking a union of predictions from the following target predictor tools: miRanda, miRDB, miRWalk, PICTAR5, RNA22 and Targetscan [112, 43, 187, 73, 111, 89, 136]. An integrative inverse-correlation analysis of miRNAs and gene expression profiles was performed to separate genuine or functional miRNA-target relationships from potentially noisy target predictions. We hypothesized that the observed differential expression of mRNAs is due to the changes in expression of their regulating miRNAs; and that the relative extent of changes in expression of miRNA targets is higher than that of the non-targets for a given miRNA.

Due to the differences in the sample sizes of relevant cell-lines in miRNA and mRNA expression datasets, normalized expression levels of each gene and C_p values for every miRNA were summarized to one value per cell line by taking an average within the relevant cell line replicate set. Since levels of functional miRNAs are expected to be inversely proportional to their mRNA targets and C_p values from qPCR are inversely proportional to expression levels, a series of tests for positive correlations between miRNA and mRNA pairs over the four resistant cell lines were performed (Figure 6.1). Such correlation analysis was performed only for 191 miRNAs that were significantly differentially expressed (miRNA logFC$\geq$0.7 and adjusted p-value$<$0.05) in any of the three tamoxifen resistance cell lines. miRNA-mRNA pairs with high positive correlation (PCC $\geq$ 0.8), implicating negative correlation between mRNA and miRNA levels, were selected for further functionality testing. These miRNA-target pairs exhibiting miRNA-mediated regulation at expression level are referred to as functional miRNA-targets in the remainder of the manuscript.

For each resistant cell line, sets of correlated miRNA-mRNA pairs were tested by two-sample Wilcoxon rank-sum tests. Briefly, each two-sided Wilcoxon test was performed on a set of fold-changes of functional miRNA targets versus fold-changes of non-targeted genes. P-values of Wilcoxon tests were corrected for multiple testing errors via the Benjamini Hochberg method [7]. If the target mRNAs had lower ranks than the background list of mRNAs, then log_{10} of Wilcoxon rank sum test p-values were negated.

Results

The miRNA expression profiles of three tamoxifen-resistant cell lines (TamR1, TamR4 and TamR8) and the parental ER-positive tamoxifen-sensitive cell line (MCF7/S0.5) were analyzed for differential expression. A number of miRNAs were significantly altered during tamoxifen-resistance, and were in common for all three tamoxifen resistant cell lines vs MCF7/S0.5 reftab:summary. In addition, a large number of miRNAs were found to exhibit altered expression in one or more of the resistant cell lines.

Table 6.1. Summary of significantly differentially expressed miRNAs and mRNAs. The numbers in the table refer to the differentially expressed miRNAs and mRNAs

miR, mRNA change	Cell line	miRNAs	miRNAs w corr. targets	mRNAs	miR-reg. mRNAs	% mRNAs under miR-reg.	Odds ratio
Up,Down	TAMR1	40	28	67	44	66	1.25
	TAMR4	51	30	231	137	59	1.23
	TAMR8	47	33	22	11	50	0.76
Down, Up	TAMR1	29	26	204	132	65	2.31
	TAMR4	52	38	205	145	71	1.23
	TAMR8	48	31	125	79	63	1.57

Supplementary table 6.2 lists significantly altered miRNAs in each of the TAMR-contrast relative to MCF7.

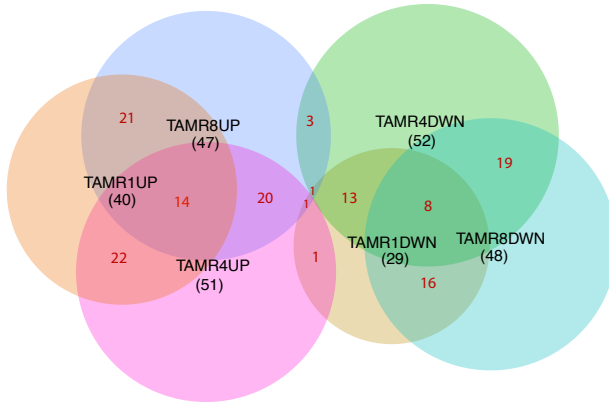

Figure 6.2. Extent of agreement in terms of miRNAs with altered expression in tamoxifen resistant cell lines. Overlap sizes represent agreement of induced and/or suppressed miRNAs during tamoxifen resistance. The circle sizes correspond to the total number of miRNAs in each set.

Three resistant cell lines were compared for their agreement in terms of their sets of miRNAs with altered expression. As seen in Figure 6.2, TAMR1 and TAMR4 have 13 repressed miRNAs in common, including oncomiR miR-95; whereas TAMR1 and TAMR8 shared 16 repressed miRNAs, including, miR-182*, miR-218, miR-23b, miR-24-1*, miR-611, miR-651, miR-519e and miR-522; while TAMR4 and TAMR8 shared 19 down-regulated miRNAs such as miR-341-3p/5p, miR-96, miR-101, miR-29b-1*, miR-183 and miR-143. In terms of miRNAs exhibiting higher expression in tamoxifen resistant vs. MCF-7/S0.5 cells, TamR1 and TamR4 shared 22 miRNAs, whereas TamR1 and TamR4 share 21 miRNAs. Among these commonly

up-regulated miRNAs were tumor suppressors miR-181b, onco-miR miR-210 and a known ER-alpha regulator miR-18a [23].

Table 6.2. miRNAs with altered expression profiles in all tamoxifen-resistant cell lines compared to MCF-7/S0.5 (adjusted p-value < 0.05).

miRNA	**LFC in cell line**		
Cell line	**TamR1**	**TamR4**	**TamR8**
miR-218	-2.92	-1.56	3.12
miR-135a	-2.84	-6.16	-4.74
miR-652	-1.99	-1.17	-1.64
miR-135b	-1.89	-2.16	-2.33
miR-95	-1.82	-2.45	-1.19
miR-1248	-1.68	-1.61	-1.01
miR-101*	-1.02	-3.05	-1.37
miR-152	1.29	1.4	1.49
miR-517c	2.06	1.66	1.06
miR-551b	2.09	5.79	3.51
miR-203	2.2	1.34	1.12
miR-519a	2.23	1.69	1.2
miR-516a-5p	3.47	2.87	2.46

Global analysis of miRNA-target relationships

Using our inverse-correlation analysis on miRNA-mRNA expression values, we derived miRNA-target pairs that, in addition to being predicted targets, also showed evidence of direct miRNA-regulation through anti-correlated expression signatures. Figure 6.2 eluciates the possible miRNA regulation event for a subset of miRNAs that were consistently up or down regulated in all tamoxifen resistant cell (Fig. 6.2 a). The first panel of the heatmap refers to the log-ratios of miRNA expression in all three TamR vs. MCF-7/S0.5 comparisons. Through the inverse-correlation analysis, we obtained a set of predicted targets of miRNAs that exhibited significant negative associations with their corresponding miRNA regulators in terms of their expression profiles. Panel II and III in the heatmap represent the extent of miRNA-regulation on such the functional targets. Direction of the fold-change for a majority of the functional miRNA targets was coherent with respect to their miRNA-regulators. Figure 6.2c shows a cumulative distribution function applied on the log-ratios of genes. Functional targets of miR-135b (and miR-516a-5p) were more up (and down) regulated compared to the non-targets

as seen by significant p-values of Wilcoxon rank sum test. P-values of the Wilcoxon rank sum test were log_{10}-transformed and assigned a sign based on direction of fold-change of target genes (panel III in Figure 6.2 b).

Next, we derived statistics to demonstrate the extent and significance of miRNA-regulated differentially expressed genes. Direction of change of expression for about 55% of the significantly differentially expressed genes (absolute log2-fold-changes ≥ 0.7, adjusted p-values ≤ 0.05) was coherent with the direction of fold-changes of their miRNA regulators.

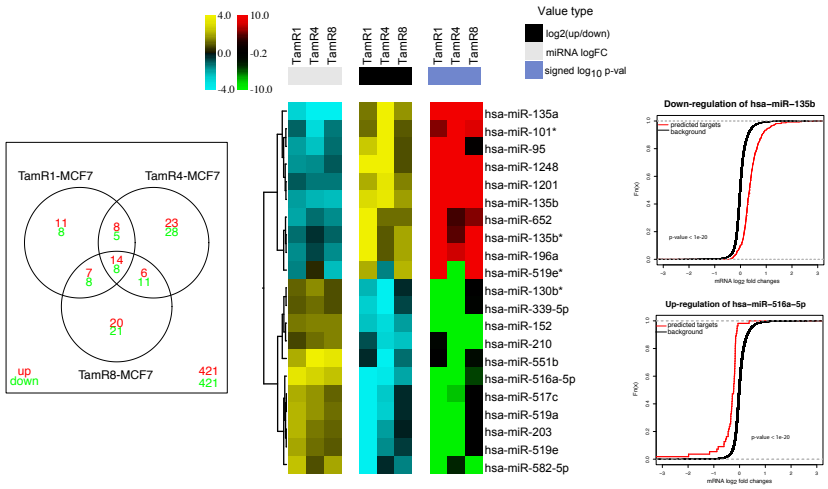

Figure 6.3. MicroRNAs with significant differential expression and expression of their functional targets. (a) Venn Diagram, number of differentially expressed in each resistant cell-line relative to MCF7. (b) Panel I of the heatmap shows miRNA log-fold changes. These miRNAs were significantly differentially expressed (absolute log2 fold-change $\geq$ 0.7 and adjusted p-value $<$ 0.05). Panel II shows (log2) ratio of counts of up- and down-regulated targets for each miRNA. (c) Signed log_{10} p-value of Wilcoxon rank-sum test.

We were particularly interested in the functions of genes under direct miRNA-mediated regulation obtained though our inverse-correlation analysis. Using genes under direct miRNA-regulation we were able to elucidate various biological modules affected by miRNA-mediated gene regulation. Such genes in TamR1 and TamR4 cell lines showed a strong enrichment (FDR $<$ 0.05 for test of over-representation) for glypican pathway, IGF1 pathway, PDGFR-beta signaling, IFN-gamma pathway, plasma membrane estrogen receptor signaling, TGF-beta signaling, nuclear and cytoplasmic SMAD2/3, BMP signaling and class I PI3K signaling events. TamR8, despite

of a milder response in terms of differential expression, showed quite good agreement with signaling events in other cell lines via a smaller sized geneset.

miRNA-regulated genes exhibiting differential expression in each of the tamoxifen resistant cell lines were investigated for their potential in elucidating molecular mechanism devised through miRNA-mediated post-transcriptional regulation. miRNAs in TamR1 were specifically targeting genes characterized for pathways in cancer (FDR < 0.05) and TGF-beta signaling (FDR < 0.05), MAPK signaling, wnt-signaling, EphrinB-EPHB and ECM-receptor signaling pathways. MicroRNAs in TamR4 seemed to target genes associated with signaling by Aurora kinases (FDR < 0.05), FOXM1 transcription factor network (FDR < 0.05), and cell cycle M Phase and G2/M transition (FDR < 0.05), Glypican pathway, FOXA transcription factor network, PDGFR-beta signaling, Kit receptor, exon guidance and PDGF signaling pathway. Although being not significant, association to functional categories for miRNA-regulated genes in TamR8 included IFN-gamma pathway, Jak-STAT signaling, plasma membrane estrogen receptor signaling and neurotrophin signaling pathways.

MicroRNA-mediated regulation of ER-alpha and PGR expression

We further examined the involvement of miRNAs on the key gene of ER-positive breast cancers, ESR1 and PGR. Our integrated analysis identified the following miRNAs as direct modulators of ER-alpha expression: miR-199a-3p, miR-582-5p, miR-519a, miR-519b-3p, miR-521, miR-522, miR-550, miR-519e, miR-220b, miR-22, miR-362-5p and miR-203. Such combinatorial regulation of ER-alpha expression is strongly evident in TamR1 cell lines, where all of the listed miRNA regulators were significantly up regulated, while TamR4 and TamR8 displayed the same pattern albeit to a lesser extent(Figure 6.4). Figure 6.4a-c depict all significantly changing miRNAs that target ESR1 and PGR, identified by inverse-correlation. Although both combinations, up and down regulated miRNAs were found, a clear and significant bias was observed for coherent changes between miRNA and their differentially expressed targets.

Gene regulation events upstream and downstream of FOXM1 in TamR4

FOXM1 appeared to be under a combinatorial targeting by number of highly expressed miRNAs as shown in Figure 6.4d. Repression of FOXM1 also correlated with the repression of its downstream targets in TamR4 but not in TamR1 or TamR8. In addition to regulating protein-coding genes, FOXM1 has also been found to regulate transcription of miR-135a [67]. Down-regulation of miR-135 is consistent with the up-regulation of its targets, EPHA4/7 in TamR4.

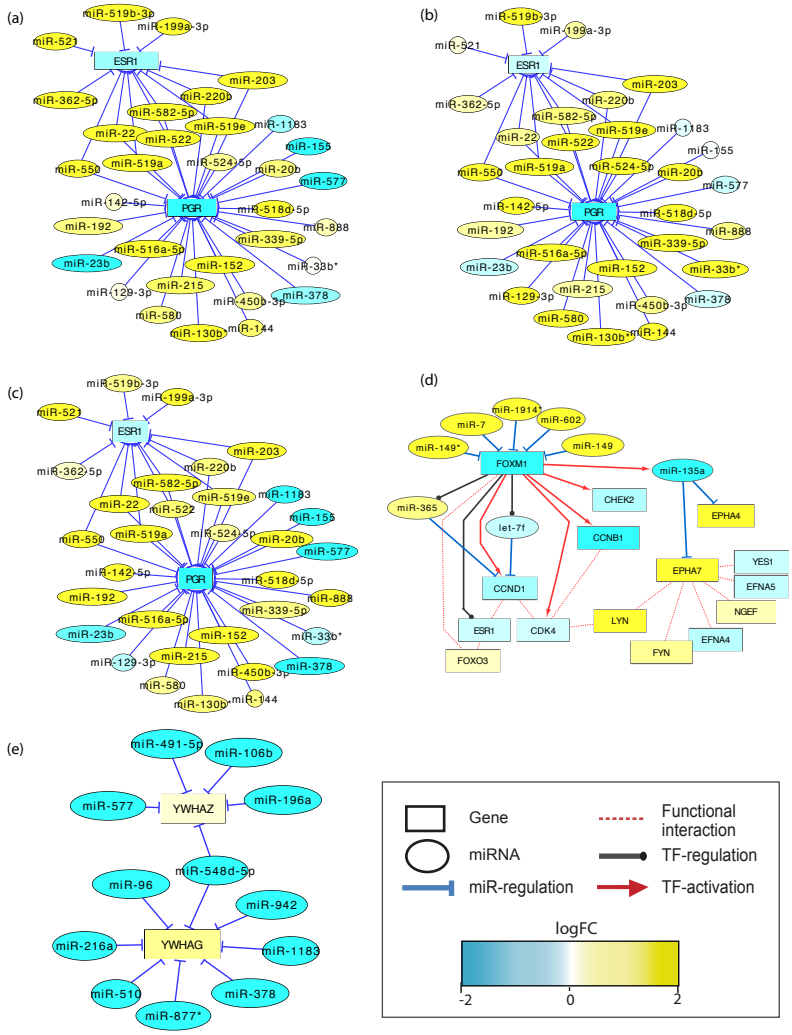

Figure 6.4. **MicroRNA-mediated gene regulation inferred from integration of differentially expressed miRNAs and their differentially expressed mRNA targets. Functional regulation of estrogen receptor alpha and progesterone receptor due to differentially expressed miRNAs in TamR1 (a), TamR4 (b) and TamR8 (c) relative to MCF-7. Evidence for FOXM1 down-regulation by number of miRNAs in TamRX is shown in D. Suppressed levels of FOXM1 are positively correlated with its transcriptional targets for cell cycle regulation. FOXM1 acts as a regulator on promoters of miR-135a, let-7f, miR-365 (d), YWHAG and YWHAZ are under direct regulation via miRNAs in TamR8 (e). Color of the node corresponds to the fold-changes observed in corresponding contrasts, whereas size of nodes is directly related to the significance of such differential expression (adjusted p-values).**

Loss of miRNA regulation for YWHAG and YWHAZ

Expression of 14-3-3 family member, YWHAZ and YWHAG, were induced in TamR8 but not in TamR1/4. These genes appeared to be under direct regulation by a relatively small number of miRNAs, as seen in Figure 6.4e. miR-96, miR-942, miR-378, miR-196a, miR-106b, miR-577, miR-491-5p showed significant down regulation in anti-correlation with YWHAG and YWHAZ in TamR8 (adjusted p-value ≤ 0.05).

Discussion

Tamoxifen resistance is a major clinical problem in treatment of ER-positive breast cancer, leading to recurrence of the disease in approximately 30% of the patients. In this study we characterized tamoxifen resistance by an extensive analysis of miRNA-mRNA expression profiles in 3 tamoxifen resistant cell lines and its parental MCF-7/S0.5 cell line. miRNAs with significantly altered expression profiles in resistant cell lines compared to MCF-7, suggests their role in drug resistance.

Based on our integrated analysis of miRNA and mRNA expression profiles, we obtained a number of functional miRNA-target relationships. Protein-coding genes often are under a direct, combinatorial regulation via multiple miRNAs, exerting a coordinate action. Such combinatorial regulatory effects have been described for various genes [195, 99, 84]. A table of genes associated with signaling pathway of tamoxifen resistance is listed with their miRNA regulators in Suppl. Table 6.2. Our analysis of miRNA-target relationships explained the extent of gene regulation under the influence of direct miRNA-mediated post-transcriptional regulation. Direction of altered expression for more than 50% of the significantly differentially expressed genes was coherent with the expression levels of their miRNA regulators. A functional enrichment analysis on miRNA-regulated genes showed their associations with many signaling pathways previously linked with tamoxifen resistance. Therefore, the miRNA-target relationships inferred by this study might be highly relevant to the acquired tamoxifen resistance.

A number of known regulatory relationships that were not identified were found to be in agreement upon further inspection. Lykkesfeldt et al. showed that a loss of expression of estrogen receptor often leads to tamoxifen resistance [122]. For example miR-22 regulates ER-alpha by direct post-transcriptional regulation [65] and although not observed in the inverse correlation analysis, expression of miR-22 and ESR1 was found to be anti-correlated with smaller fold-changes. MiR-221 and miR-222, regulators of ER-alpha expression were studied under both tamoxifen sensitive (MCF7) and tamoxifen resistant cell lines suggest that over expression of these miRNAs is correlated to the drug resistance [134]. Our data supported this hypothesis only partially. MiR-221 expression was slightly induced in TamR1 and TamR4 but not in TamR8, whereas expression of miR-222 was suppressed specifically in TamR4. Given that miR-221/222 share identical

seed sequence and are located less than 1kb apart on chromosome X, an upstream regulation is more likely to modulate miR-222 levels than an epigenetic effect.

Our systematic approach of integrating various evidences of gene regulation with miRNA-target expression profiles revealed a small linear pathway in TamR4. Forkhead box protein, FOXM1, also a known regulator of estrogen receptor alpha in breast cancer cells [120], is associated with positive regulation of its direct transcriptional targets CCNB1, CCNA2, PLK1, aurora B kinase, Cdc25B phosphatase and CENPF. FOXM1 is a modulator of G1/S and G2/M transition, cell cycle, DNA replication, chromosomal segregation and stability during mitosis. FOXM1 has been recently characterized for binding to the promoter of miR-135a and exerting a positive regulatory effect on miR-135a [120]. Among all resistant cell lines, miR-135a levels are most down regulated in TamR4, suggesting an additional regulation of miR-135a transcription through FOXM1 (Figure 6.4),especially in TamR4. MiR-135 modulates expression of several genes in TamR4, among which EPHA4 and EPHA7 are specifically up regulated in TamR4 but not in other resistant cell lines. Due to the functional interactions of these genes with FYN, YES1, LYN, etc, Ephrins are beeing seen as therapeutic targets in cancer due to their crucial role in development and progression of cancer, metastasis as well as in the inhibition of tumorigenicity [16]. Also, high expression levels of EphA4 and EphA7 have recently been associated with poor overall survival in breast cancer [44]. Taken together, FOXM1 desregulation, followed by a loss of function of miR-135a and a loss of regulation of Eph receptors may play role in development of drug resistance in TamR4. An experimental validation of protein expression of FOXM1, ephrins as well as a silencing experiment for miR-135a is required to confirm this hypothesis.

The 14-3-3 (YWHA) proteins are highly conserved proteins that have been implicated in regulating cell cycle, apoptosis and recently also in cancers. 14-3-3γ (YWHAG) has been shown to exert its oncogenic effects through activation of MAPK and PIK pathways [144]. 14-3-3ζ or YWHAGZ is often up regulated in patients receiving tamoxifen adjuvant therapy. In addition, over expression of this gene is related to poor patient survival [8, 58, 59]. In our dataset, both of these genes were up regulated in TamR8 but not in TamR1/4. miR-96, miR-942, miR-378, miR-196a, miR-106b, miR-577 and miR-491-5p showed significant down regulation in TamR8 (adjusted p-value *leq* 0.05), expressing a direct effect of miRNA-mediated regulation. Bergamaschi et al [9], recently showed that tamoxifen bears a suppressive role on expression of miR-451 in on TamR cell lines by inducing loss-of-regulation of YWHAZ due to down-regulation of miR-451, its miRNA regulator. We, however, did not find a significant over expression of miR-451 in our dataset. This is likely due to the fact that miR-451 was flagged a as poor quality assay in our dataset due to multiple peaks observed in the melting curve characteristics of its qPCR primer. Nevertheless, this small yet critical event

of miRNA-mediated regulation of 14-3-3 family member genes might be of high clinical relevance after further experimental validations.

miRNAs with altered expression in tamoxifen resistant cell lines showed an agreement with the findings of earlier studies within the area of tamoxifen resistance. Lyng et al [125] performed a study on a large cohort of 152 primary breast cancer patients treated with tamoxifen, grouped by their recurrence status and various other clinical characteristics. We tried to investigate whether the differences in the miRNA expression signatures in the resistant cell lines showed any agreement with the differences in the expression patterns observed between recurrent and nonrecurrent groups. Over-expression of miR-205 and miR-363 in TamR8 corresponds to over expression seen in the recurrent group in test set # 1 that consisted of samples of recurrent type and characterized by a high levels of tumor infiltrated lymphnodes (8.6, on an average). Significant down-regulation of miR-29b in TamR4 relates to a dysregulation of this miRNA in testset # 3, characterized by 3-5 years of average tamoxifen therapy. Further, miR-190b and miR-135a, down-regulated in all tamoxifen resistant cell lines were also the most significantly dysregulated miRNAs between recurrent and non-recurrent groups in the discovery set that consisted of a patient-group with an average 2 years of exposure to tamoxifen therapy, and a positive lymph nodes status. Therefore, a systematic comparison of the findings achieved through the use of cell line model to clinical samples would help explain the mechanisms of resistance *in vivo* .

Next, we assessed whether the observed miRNA expression changes were in agreement with the already known tamoxifen resistance related miRNAs. Over expression of miR-181b expression has previously been studied with a view of tamoxifen resistance [121]. In our dataset, miR-181b was consistently up regulated in all resistant cell lines, suggesting its contribution to a common machinery to confer drug resistance. A closer look at the functional targets of miR-181b provided us with set of key regulators of breast cancer and drug resistance related genes with significant down regulation: for example, HEY1, CA2, PIK3R1, LYN, ESR1, JUN, STAT1, MYB, BCL2, CYCS, BAMBI, CTGF, and SOX9. Another miRNA consistently down regulated in all resistant cell lines was miR-342. Down regulation of miR-342 has been linked with tamoxifen resistance by Cittelly et al [32]. We identified 14 functional targets of miR-342-3p and -5p, some of which were FYN, TGFBR1, COL4A6, CDKN1A, and Ephrins EPHA4/7 (Suppl. Fig 6.5). Taken together, targets of previously characterized miR-181b and miR-342 were highly relevant to anti-estrogen resistance and therefore, could be investigated for their potential in restoration of drug sensitivity.

Conclusion

In conclusion, our systematic approach of inferring miRNA-target relationships elucidated important and detailed lines of events contributing to the development of drug resistance. Several miRNA-target relationships were identified, many of which showed significant association to various signaling events known to be associated with the development of tamoxifen resistance. Combinatorial miRNA targeting as well as post-transcriptional modifications to the upstream promoters of miRNAs and/or their targets. Therefore, an experimental validations of some of our findings is necessary and is in progress. Nonetheless, this study provides hints toward such events occurring during tamoxifen resistance and is likely to correspond with the miRNA-mediated gene regulation in patient material of specific characteristics.

Supplementary files

Table 6.3. List of significantly differentially expressed miRNAs. miRNAs with significantly altered expression profiles in tamoxifen resistant cell lines relative to MCF7 cell line. (Absolute log-ratio of expression $\geq$1 AND adjusted p-value $\leq$ 0.05)

Contrast (relative to MCF7 cell line)	List of differentially expressed miRNAs
TAMR1	hsa-miR-101*, hsa-miR-1203, hsa-miR-1248, hsa-miR-130b*, hsa-miR-132, hsa-miR-135a, hsa-miR-135b, hsa-miR-142-3p, hsa-miR-152, hsa-miR-153, hsa-miR-182*, hsa-miR-1913, hsa-miR-196a, hsa-miR-196b, hsa-miR-1979, hsa-miR-199b-5p, hsa-miR-200a*, hsa-miR-203, hsa-miR-212, hsa-miR-215, hsa-miR-218, hsa-miR-22, hsa-miR-220b, hsa-miR-23b, hsa-miR-24-1*, hsa-miR-339-5p, hsa-miR-33b, hsa-miR-345, hsa-miR-34b, hsa-miR-362-5p, hsa-miR-516a-5p, hsa-miR-517b, hsa-miR-517c, hsa-miR-519a, hsa-miR-519b-3p, hsa-miR-519e, hsa-miR-519e*, hsa-miR-522, hsa-miR-551b, hsa-miR-577, hsa-miR-582-5p, hsa-miR-593, hsa-miR-611, hsa-miR-649, hsa-miR-652, hsa-miR-943, hsa-miR-95
TAMR4	hsa-miR-1, hsa-miR-101„ hsa-miR-101*, hsa-miR-1201, hsa-miR-1248, hsa-miR-130b*, hsa-miR-135a, hsa-miR-135b, hsa-miR-138, hsa-miR-143, hsa-miR-149, hsa-miR-149*, hsa-miR-152, hsa-miR-153, hsa-miR-181b, hsa-miR-181c, hsa-miR-183, hsa-miR-18a, hsa-miR-18b, hsa-miR-190, hsa-miR-1908, hsa-miR-190b, hsa-miR-191*, hsa-miR-192*, hsa-miR-193a-3p, hsa-miR-1979, hsa-miR-200a, hsa-miR-203, hsa-miR-21, hsa-miR-210, hsa-miR-2113, hsa-miR-218, hsa-miR-26a-2*, hsa-miR-29b, hsa-miR-30a, hsa-miR-330-3p, hsa-miR-337-3p, hsa-miR-339-5p, hsa-miR-33a*, hsa-miR-33b, hsa-miR-33b*, hsa-miR-342-3p, hsa-miR-342-5p, hsa-miR-345, hsa-miR-424, hsa-miR-424*, hsa-miR-499-5p, hsa-miR-503, hsa-miR-516a-5p, hsa-miR-517c, hsa-miR-519a, hsa-miR-519e, hsa-miR-520d-3p, hsa-miR-522, hsa-miR-524-5p, hsa-miR-548k, hsa-miR-551b, hsa-miR-580, hsa-miR-600, hsa-miR-628-5p, hsa-miR-629*, hsa-miR-652, hsa-miR-7, hsa-miR-708, hsa-miR-720, hsa-miR-941, hsa-miR-95, hsa-miR-96
TAMR8	hsa-let-7i, hsa-miR-101, hsa-miR-101*, hsa-miR-106b, hsa-miR-1201, hsa-miR-1203, hsa-miR-1248, hsa-miR-129*, hsa-miR-135a, hsa-miR-135b, hsa-miR-142-3p, hsa-miR-143, hsa-miR-152, hsa-miR-182*, hsa-miR-183, hsa-miR-1908, hsa-miR-190b, hsa-miR-192, hsa-miR-194, hsa-miR-196a, hsa-miR-203, hsa-miR-205, hsa-miR-20b, hsa-miR-210, hsa-miR-215, hsa-miR-218, hsa-miR-22, hsa-miR-23b*, hsa-miR-24-1*, hsa-miR-24-2*, hsa-miR-30a, hsa-miR-328, hsa-miR-340, hsa-miR-342-5p, hsa-miR-363, hsa-miR-375, hsa-miR-378, hsa-miR-378*, hsa-miR-449a, hsa-miR-455-3p, hsa-miR-455-5p, hsa-miR-491-5p, hsa-miR-503, hsa-miR-516a-5p, hsa-miR-517c, hsa-miR-519a, hsa-miR-519e*, hsa-miR-551b, hsa-miR-574-3p, hsa-miR-577, hsa-miR-582-5p, hsa-miR-590-5p, hsa-miR-611, hsa-miR-628-3p, hsa-miR-652, hsa-miR-720, hsa-miR-877*, hsa-miR-93*, hsa-miR-942, hsa-miR-95, hsa-miR-96

Table 6.4: **Functional miRNA-target relationship inferred by inverse-correlation analysis**

Gene	Regulating miRNAs
MYCN	hsa-miR-30a, hsa-miR-34b*, hsa-miR-449a
NFKB1	hsa-miR-548d-5p, hsa-miR-577
YWHAG	hsa-miR-1183, hsa-miR-216a, hsa-miR-378, hsa-miR-510, hsa-miR-548d-5p, hsa-miR-877*, hsa-miR-942, hsa-miR-96
YWHAZ	hsa-miR-106b, hsa-miR-196a, hsa-miR-491-5p, hsa-miR-548d-5p, hsa-miR-577
FOXM1	hsa-miR-149, hsa-miR-149*, hsa-miR-1914*, hsa-miR-602, hsa-miR-7
BMP7	hsa-miR-199a-5p
CD59	hsa-miR-1201, hsa-miR-1248, hsa-miR-135b, hsa-miR-190b, hsa-miR-196a, hsa-miR-23b, hsa-miR-24-1*, hsa-miR-378*, hsa-miR-429, hsa-miR-485-3p, hsa-miR-577, hsa-miR-593, hsa-miR-616*, hsa-miR-627
CDKN2B	hsa-miR-1244, hsa-miR-155, hsa-miR-196a, hsa-miR-23b, hsa-miR-429, hsa-miR-519e*, hsa-miR-577, hsa-miR-93*
CLU	hsa-miR-1913, hsa-miR-378
COL5A1	hsa-miR-199a-5p, hsa-miR-218, hsa-miR-517b, hsa-miR-649, hsa-miR-766
CPE	hsa-miR-335*, hsa-miR-33b, hsa-miR-616*
DUSP6	hsa-miR-1248, hsa-miR-33b, hsa-miR-649
EFNB2	hsa-miR-153, hsa-miR-199a-5p, hsa-miR-218, hsa-miR-33b, hsa-miR-34b, hsa-miR-548i, hsa-miR-767-5p
EPAS1	hsa-miR-196a, hsa-miR-23b, hsa-miR-577
FZD7	hsa-miR-149*, hsa-miR-34b, hsa-miR-593, hsa-miR-766
FZD8	hsa-miR-517b, hsa-miR-943
IRF6	hsa-miR-1248, hsa-miR-23b
ITPR1	hsa-miR-429, hsa-miR-577
LYN	hsa-miR-1248, hsa-miR-147b, hsa-miR-199a-5p, hsa-miR-33b, hsa-miR-513a-3p
MAGED1	hsa-miR-513a-3p
MAP1B	hsa-miR-1248, hsa-miR-135b*, hsa-miR-153, hsa-miR-23b, hsa-miR-33b, hsa-miR-593, hsa-miR-616*, hsa-miR-649, hsa-miR-943
MYO1B	hsa-miR-155, hsa-miR-378, hsa-miR-548d-5p, hsa-miR-611
NOG	hsa-miR-429, hsa-miR-593
NRP1	hsa-miR-1248, hsa-miR-153, hsa-miR-23b, hsa-miR-485-3p, hsa-miR-604
PCDH8	hsa-miR-101, hsa-miR-153
PLAU	hsa-miR-23b, hsa-miR-593
PRICKLE2	hsa-miR-1248, hsa-miR-135b, hsa-miR-153, hsa-miR-182*, hsa-miR-196a, hsa-miR-23b, hsa-miR-33b, hsa-miR-593, hsa-miR-616*, hsa-miR-627, hsa-miR-649
PRKCA	hsa-miR-1248, hsa-miR-135b, hsa-miR-135b*, hsa-miR-147b, hsa-miR-182*, hsa-miR-183*, hsa-miR-196a, hsa-miR-23b, hsa-miR-30a*, hsa-miR-485-3p, hsa-miR-548k, hsa-miR-604, hsa-miR-616*, hsa-miR-95
SDC2	hsa-miR-101, hsa-miR-1183, hsa-miR-135b, hsa-miR-182*, hsa-miR-23b, hsa-miR-24-1*, hsa-miR-485-3p, hsa-miR-577, hsa-miR-627
SH3BGRL	hsa-miR-135b, hsa-miR-196a, hsa-miR-485-3p, hsa-miR-649
TCF7L1	hsa-miR-218, hsa-miR-455-3p, hsa-miR-517*
TGFB2	hsa-miR-593
THBS1	hsa-miR-1913

ARHGEF6	hsa-miR-1253, hsa-miR-130b*, hsa-miR-144, hsa-miR-199b-5p, hsa-miR-220b, hsa-miR-33a*, hsa-miR-522, hsa-miR-582-5p, hsa-miR-589*
CACNA1D	hsa-miR-142-3p, hsa-miR-144, hsa-miR-199b-5p, hsa-miR-519e, hsa-miR-522, hsa-miR-582-5p
CAP2	hsa-miR-132, hsa-miR-203, hsa-miR-212, hsa-miR-522, hsa-miR-582-5p
ELOVL2	hsa-miR-144, hsa-miR-152, hsa-miR-18a, hsa-miR-203, hsa-miR-22, hsa-miR-339-5p, hsa-miR-519e, hsa-miR-550, hsa-miR-888
ERBB4	hsa-miR-1253, hsa-miR-130b*, hsa-miR-144, hsa-miR-18a, hsa-miR-18b, hsa-miR-199b-5p, hsa-miR-203, hsa-miR-22, hsa-miR-220b, hsa-miR-339-5p, hsa-miR-33a*, hsa-miR-345, hsa-miR-362-5p, hsa-miR-518d-5p, hsa-miR-518e*, hsa-miR-519a, hsa-miR-519b-3p, hsa-miR-519e, hsa-miR-522, hsa-miR-524-5p, hsa-miR-525-5p, hsa-miR-580, hsa-miR-589*, hsa-miR-609
ESR1	hsa-miR-199a-3p, hsa-miR-203, hsa-miR-22, hsa-miR-220b, hsa-miR-362-5p, hsa-miR-519a, hsa-miR-519b-3p, hsa-miR-519e, hsa-miR-521, hsa-miR-522, hsa-miR-550, hsa-miR-582-5p
GREB1	hsa-miR-130b*, hsa-miR-152, hsa-miR-192, hsa-miR-203, hsa-miR-20b, hsa-miR-215, hsa-miR-22, hsa-miR-220b, hsa-miR-550, hsa-miR-582-5p
GULP1	hsa-miR-144, hsa-miR-199a-3p, hsa-miR-203, hsa-miR-33a*, hsa-miR-362-5p, hsa-miR-522, hsa-miR-582-5p
KIAA1324	hsa-miR-519e
MGP	hsa-miR-152
NFKBIA	hsa-miR-519a
NPY1R	hsa-miR-1253, hsa-miR-132, hsa-miR-203, hsa-miR-212, hsa-miR-522, hsa-miR-582-5p
SGK1	hsa-miR-142-3p, hsa-miR-144, hsa-miR-522, hsa-miR-582-5p
SULF1	hsa-miR-1, hsa-miR-138, hsa-miR-196b, hsa-miR-330-3p, hsa-miR-330-5p, hsa-miR-340, hsa-miR-34b*
TM4SF1	hsa-miR-491-5p, hsa-miR-548d-5p, hsa-miR-640
ACADL	hsa-miR-135a, hsa-miR-135b, hsa-miR-32, hsa-miR-424, hsa-miR-503, hsa-miR-548k
ARMCX1	hsa-miR-424, hsa-miR-503, hsa-miR-96
ATP9A	hsa-miR-1248, hsa-miR-135a, hsa-miR-135b, hsa-miR-187, hsa-miR-190b, hsa-miR-30a*, hsa-miR-33b, hsa-miR-342-5p, hsa-miR-548k, hsa-miR-604, hsa-miR-627, hsa-miR-649, hsa-miR-708
CAMK2B	hsa-miR-138, hsa-miR-342-3p, hsa-miR-708*
CD36	hsa-miR-143, hsa-miR-216a, hsa-miR-486-5p, hsa-miR-96
CPT1C	hsa-miR-342-3p
CTGF	hsa-miR-138, hsa-miR-181c, hsa-miR-330-3p, hsa-miR-330-5p, hsa-miR-597, hsa-miR-708*
DDX58	hsa-miR-101, hsa-miR-135a, hsa-miR-181c, hsa-miR-193a-3p, hsa-miR-30a*, hsa-miR-32, hsa-miR-330-5p, hsa-miR-335*, hsa-miR-597, hsa-miR-600, hsa-miR-708*
EGR2	hsa-miR-138, hsa-miR-181c, hsa-miR-190, hsa-miR-200a, hsa-miR-330-3p, hsa-miR-335*, hsa-miR-424*, hsa-miR-548i, hsa-miR-548k, hsa-miR-708
EPHA4	hsa-miR-1201, hsa-miR-1248, hsa-miR-135a, hsa-miR-135b, hsa-miR-181c, hsa-miR-183*, hsa-miR-193a-3p, hsa-miR-335*, hsa-miR-342-3p, hsa-miR-342-5p, hsa-miR-486-5p, hsa-miR-548k, hsa-miR-627, hsa-miR-708
FDFT1	hsa-miR-181c, hsa-miR-424, hsa-miR-548k, hsa-miR-708
FYN	hsa-miR-33b, hsa-miR-342-5p
IFIT1	hsa-miR-183*, hsa-miR-600
IFNB1	hsa-miR-138
IFNGR1	hsa-miR-183*, hsa-miR-335*

IL15	hsa-miR-101, hsa-miR-101*, hsa-miR-200a, hsa-miR-32, hsa-miR-424
IRF9	hsa-miR-302e, hsa-miR-600, hsa-miR-627
JUP	hsa-miR-342-5p, hsa-miR-424, hsa-miR-627, hsa-miR-708
OASL	hsa-miR-15a*, hsa-miR-183*, hsa-miR-193a-3p, hsa-miR-338-3p, hsa-miR-597, hsa-miR-708
PLCXD3	hsa-miR-153, hsa-miR-181c, hsa-miR-183*, hsa-miR-190, hsa-miR-199a-5p, hsa-miR-26a-1*, hsa-miR-302e, hsa-miR-30a, hsa-miR-330-3p, hsa-miR-335*, hsa-miR-338-3p, hsa-miR-33b, hsa-miR-342-3p, hsa-miR-499-5p, hsa-miR-513a-3p, hsa-miR-548i, hsa-miR-548k, hsa-miR-597
PRKD1	hsa-miR-200a
PTGES	hsa-miR-193a-3p, hsa-miR-330-3p
RAPGEFL1	hsa-miR-138, hsa-miR-302e, hsa-miR-337-3p, hsa-miR-424, hsa-miR-486-5p, hsa-miR-520d-3p, hsa-miR-604, hsa-miR-708
RPRM	hsa-miR-105*, hsa-miR-26a-1*, hsa-miR-30a*, hsa-miR-33b, hsa-miR-424, hsa-miR-424*, hsa-miR-627
STAT1	hsa-miR-101, hsa-miR-1248, hsa-miR-135a, hsa-miR-135b, hsa-miR-181c, hsa-miR-183*, hsa-miR-335*, hsa-miR-548k
TGFBR3	hsa-miR-1, hsa-miR-138, hsa-miR-181c, hsa-miR-193a-3p, hsa-miR-200a, hsa-miR-330-3p, hsa-miR-330-5p, hsa-miR-335*, hsa-miR-338-3p, hsa-miR-342-3p, hsa-miR-342-5p, hsa-miR-424, hsa-miR-628-5p
TLR2	hsa-miR-548k
TNFAIP3	hsa-miR-1, hsa-miR-101, hsa-miR-1201, hsa-miR-138, hsa-miR-15a*, hsa-miR-183*, hsa-miR-193a-3p, hsa-miR-200a, hsa-miR-330-5p, hsa-miR-335*, hsa-miR-337-3p, hsa-miR-424, hsa-miR-503
ADIPOR2	hsa-miR-223, hsa-miR-524-5p, hsa-miR-525-5p, hsa-miR-629*
AURKA	hsa-miR-129-3p, hsa-miR-149, hsa-miR-149*, hsa-miR-629*
BAMBI	hsa-miR-142-5p, hsa-miR-20b, hsa-miR-888
BCL2L12	hsa-miR-7
C12orf32	hsa-miR-18a, hsa-miR-18b
CCNA2	hsa-miR-1267, hsa-miR-524-5p, hsa-miR-589*
CCNB1	hsa-miR-144
CCND1	hsa-miR-149, hsa-miR-18b, hsa-miR-1914, hsa-miR-524-5p, hsa-miR-593*, hsa-miR-602
CDCA8	hsa-miR-149*
CDKN1B	hsa-miR-1267, hsa-miR-223, hsa-miR-33b*, hsa-miR-524-5p, hsa-miR-580
CENPA	hsa-miR-524-5p, hsa-miR-629*
CENPF	hsa-miR-33b*
CTNNAL1	hsa-miR-525-5p
CXCL12	hsa-miR-137, hsa-miR-152, hsa-miR-20b, hsa-miR-455-5p, hsa-miR-518d-5p, hsa-miR-550, hsa-miR-609, hsa-miR-720, hsa-miR-888
EIF4EBP1	hsa-miR-18a, hsa-miR-551b, hsa-miR-609
FOXC1	hsa-miR-149, hsa-miR-1914*, hsa-miR-345, hsa-miR-517c, hsa-miR-518d-5p, hsa-miR-518e*, hsa-miR-522, hsa-miR-524-5p, hsa-miR-609, hsa-miR-639
GAB2	hsa-miR-1267, hsa-miR-149, hsa-miR-152, hsa-miR-18a, hsa-miR-18b, hsa-miR-1908, hsa-miR-192*, hsa-miR-20b, hsa-miR-339-5p, hsa-miR-518d-5p, hsa-miR-550, hsa-miR-886-3p
GNB4	hsa-miR-142-5p, hsa-miR-192*, hsa-miR-20b, hsa-miR-450b-3p, hsa-miR-455-5p, hsa-miR-517*, hsa-miR-518d-5p, hsa-miR-888
IGF1R	hsa-let-7i, hsa-miR-142-5p, hsa-miR-152, hsa-miR-20b, hsa-miR-375, hsa-miR-450b-3p, hsa-miR-455-5p

IL6ST	hsa-miR-223
IRS2	hsa-miR-137, hsa-miR-142-5p, hsa-miR-152, hsa-miR-192, hsa-miR-192*, hsa-miR-215, hsa-miR-375, hsa-miR-550
KIF23	hsa-miR-191*, hsa-miR-223, hsa-miR-7
NCAPD2	hsa-miR-33b*, hsa-miR-519e*
NCAPG	hsa-miR-629*
NCAPH	hsa-miR-1267, hsa-miR-149*, hsa-miR-593*
NEK2	hsa-miR-629*
PCDH7	hsa-miR-142-5p, hsa-miR-144, hsa-miR-524-5p, hsa-miR-580
PCK2	hsa-miR-18a, hsa-miR-223, hsa-miR-550, hsa-miR-580
PPM1E	hsa-miR-1256, hsa-miR-137, hsa-miR-1908, hsa-miR-205, hsa-miR-20b, hsa-miR-888
PRKAR2B	hsa-miR-1256, hsa-miR-130b*, hsa-miR-137, hsa-miR-19b-2*, hsa-miR-450b-3p, hsa-miR-518d-5p, hsa-miR-524-5p, hsa-miR-580
RAB11FIP3	hsa-miR-149, hsa-miR-941
RAB31	hsa-miR-1256, hsa-miR-130b*, hsa-miR-149, hsa-miR-152, hsa-miR-339-5p, hsa-miR-33a*, hsa-miR-518d-5p, hsa-miR-524-5p, hsa-miR-580, hsa-miR-589*
RACGAP1	hsa-miR-130b*, hsa-miR-149, hsa-miR-223
SMC2	hsa-miR-142-5p, hsa-miR-144, hsa-miR-223, hsa-miR-524-5p, hsa-miR-525-5p, hsa-miR-580, hsa-miR-7
SOX3	hsa-miR-22, hsa-miR-518d-5p
SVIL	hsa-miR-524-5p, hsa-miR-525-5p, hsa-miR-580, hsa-miR-589*
TFAP2C	hsa-miR-1267, hsa-miR-550, hsa-miR-602
TLE1	hsa-miR-191*, hsa-miR-524-5p
TSPAN5	hsa-miR-144, hsa-miR-199a-3p
CHFR	hsa-miR-1913, hsa-miR-577, hsa-miR-593
CP	hsa-miR-548d-5p
CROT	hsa-miR-106b, hsa-miR-1183, hsa-miR-590-5p, hsa-miR-611
IL7	hsa-miR-548d-5p
INPP4B	hsa-miR-216a, hsa-miR-590-5p, hsa-miR-640
IRX2	hsa-miR-93*, hsa-miR-942
LEPR	hsa-miR-216a, hsa-miR-378, hsa-miR-510, hsa-miR-584, hsa-miR-590-5p
PDCD6	hsa-miR-548d-5p, hsa-miR-590-5p
PRLR	hsa-miR-106b, hsa-miR-548d-5p
TPD52	hsa-miR-135b*, hsa-miR-155, hsa-miR-182*, hsa-miR-196a, hsa-miR-23b, hsa-miR-429, hsa-miR-548d-5p
BMPR1B	hsa-miR-205, hsa-miR-20b, hsa-miR-218, hsa-miR-328, hsa-miR-375, hsa-miR-554, hsa-miR-628-3p, hsa-miR-888
OSMR	hsa-miR-513a-3p, hsa-miR-943

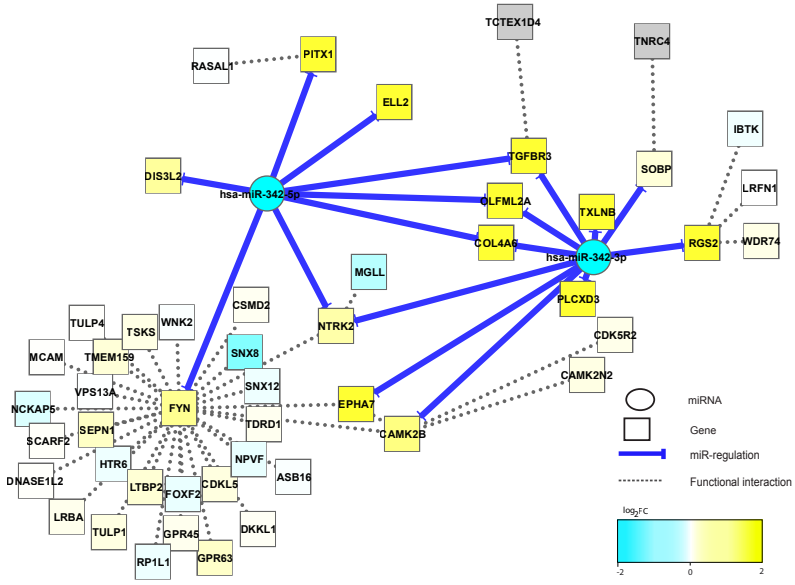

Figure 6.5. Regulation of FYN through miR-342-5p and miR-342-3p. Solid blue edges indicate miRNA regulation at post-transcription level. Dotted edges represent protein-protein interactions. FYN is a direct target of miR-342-5p.

Chapter 7

MicroRNAs in Disease Progression

7.1 Prelude

The project presented as the manuscript *"MicroRNA profiling suggest a role for MYC and NFKB1 mediated dysregulation of microRNA expression in the transformation of low-grade ocular MALT to diffuse large B-cell lymphoma"'* aims at characterizing the role of miRNAs in the transformation of low-grade ocular mucosa associated lymphoid tissue lymphoma (MALT) to a high-grade and malignant diffuse large B-cell lymphoma (DLBCL) using miRNA expression profiling on patient samples.

MALT, although being an indolent and relatively rare disease, the mechanisms by which the disease transforms into the malignant lymphoma are not very well known. A recent study with the goal of understanding such transformation for gastric MALT to gastric DLBCL identified the signature of dysregulated miRNAs and their interplay with MYC, an oncogene [38]. However, the molecular aspects of the transformation from ocular MALT to ocular DLBCL have not explored in much detail due to the limited amount of diagnostic material and also the need for frozen tissue samples. MicroRNAs, on the other hand can still be preserved in the formalin fixed paraffin embedded (FFPE) samples and very small amounts of materials are needed for the analysis. Therefore, more than 40 patient samples of ocular MALT and ocular DLBCL were analyzed for miRNA expression using microarrays. This project was a collaboration with the Department of Hematology and the Department of Pathology, Rigshospitalet, Copenhagen, Denmark. miRNA microarrays and qPCR profiling was performed at Rigshospitalet.

7.2 Manuscript

MicroRNA profiling suggest a role for MYC and NFKB1 mediated dysregulation of microRNA expression in the transformation of low-grade ocular MALT to diffuse large B-cell lymphoma

Christoffer Hother[1*], Peter Kristian Rasmussen[2*], Tejal Joshi[3], Ditte Reker[1], Ulrik Ralfkiaer[1], Christopher T. Workman[3], Steffan Heegaard[2], Elisabeth Ralfkiær[2], Kirsten Grønbæk[1†]

1 Department of Hematology, Rigshospitalet, DK-2100 Copenhagen Ø, Denmark
2 Department of Pathology, Rigshospitalet, DK-2100 Copenhagen Ø, Denmark
3 Center for Biological Sequence Analysis, Technical University of Denmark, Kgs. Lyngby, Denmark
∗ These authors contributed equally to this manuscript.
† Corresponding author. Email: kirsten.groenbaek@rh.regionh.dk

Keywords: ER-positive, Tamoxifen resistance, miRNAs, breast cancer

Abstract

Ocular adnexal lymphoma, although rare, are one of the most common type kind of malignancies involving orbital adnexal regions of the eye. Often indolent in nature, the low-grade mucosa associated lymphoid tissue lymphoma (MALT) can transform into the malignant diffuse large B cell lymphomas (DLBCLs) of eye. It is unclear, however, that which genetic events lead to the transformation of low-grade disease to a malignant and potentially fatal disease. We performed a global miRNA expression profiling of 18 MALT and 25 DLBCL clinical samples to investigate changes in the miRNA expression associated with the disease transformation. Our data analysis revealed 40 miRNAs with altered expression profiles in DLBCL with respect to MALT. Twenty-nine of the down regulated miRNAs showed enrichment for a direct transcriptional repression by MYC, an oncoprotein. We also report a possible loss-of-regulation of NFKB1 and its downstream miRNAs.

Introduction

Ocular adnexal lymphoma, found in the eyelid, conjunctiva, lacrimal sac, lacrimal gland, and orbit, is a heterogeneous group of malignancies representing 6-8% of the extranodal non-Hodgkin lymphomas (NHL) [189, 60] and, although rare, is the most common malignancy of the orbit [168]. Mucosa-associated lymphoid tissue lymphoma (MALT), also known extranodal marginal zone lymphoma or EMZL, is the most frequent ocular lymphoma and is indolent in nature [169]. A prolonged history of ocular MALT (eMALT) is often followed by disease progression and transformation toward diffuse large B-cell lymphoma in the orbit or adnexa (eDLBCL). Further, it has recently been shown that during a 52 months follow-up, 4% of the low-grade eMALT patient samples transformed to diffuse large B cell lymphoma in a large cohort of OAL patients [155]. However, it is unknown how many DLBCLs are indeed preceded by a MALT lesion, suggesting that the numbers may be higher. The molecular aspects distinguishing ocular MALT from ocular DLBCL have not been explored in much detail, mainly due to the limited amount of diagnostic material and the requirement of fresh frozen tissue for most diagnostic methods. Most previous studies of tumor suppressor pathways in lymphoma have focused on the disruption of the coding sequences of genes. However, it is now evident that epigenetic alterations and aberrant expression of non-coding RNAs, such as microRNAs, may be equally important for disease development and transformation to malignant forms of disease.

MicroRNAs (miRNAs) are 19-22 nucleotide long noncoding RNAs that target complementary regions in the 3'UTR of protein encoding mRNAs and may disrupt gene expression by either mRNA degradation or translational inhibition. A single miRNA may influence the expression of number of protein encoding genes, therefore, even moderate changes in miRNA expression may have great impact on tumorigenesis [37, 163]. For example, miR-16, miR-143, miR-145 and let-7 family miRNAs have tumor suppressive effects; whereas miR-125b-1, miR-1 and miR-17-92 cluster members are linked with tumor proliferative or oncogenic activities [49]. Dysregulation of miRNAs, whether from genetic or epigenetic events, can lead to the loss of miRNA regulation (loss-of-function) [160] and often disrupt the expression of protein coding genes [6]. Recent work by Craig et al. [38] describes the interplay between MYC, miR-34a and FOXP1 during transformation of gastric MALT to gastric DLBCL.

In the present study we attempt to distinguish the differences in miRNA regulation between ocular MALT and ocular DLBCL using miRNA expression profiling of formalin-fixed paraffin-embedded (FFPE) specimens from patients suffering from these diseases. We aimed to understand mechanisms for low-grade eMALT to eDLBCL transformation by applying an integrative bioinformatics approach to miRNA targeting networks and miRNA expres-

sion profiles.

Material and methods

Patient samples

The study includes specimens from 25 patients with ocular DLBCL, and 18 patients with ocular MALT. All eMALTs were orbital biopsies, while the 25 DLBCLs were biopsied from the orbit (18), lacrimal gland (3), eyelid (3) and conjunctiva (1). The FFPE specimens were retrieved from several Danish pathological departments during year 1980 to year 2009. For histopathological examination, all sections were stained with haematoxylin-eosin (H&E) and analysed immunohistochemically using the following panel of antibodies: bcl-2, bcl-6, CD3, CD5, CD10, CD20, CD23, CD 79α, cyclin D-1, MUM-1 and MIB-1 (Ki-67). Two independent pathologists (ER, SH) examined and reviewed the samples and, in agreement, reclassified the specimens according to the World Health Organization Classification of Tumours of Haematopoietic and Lymphoid Tissues [22]. As per classification by Hans et al. [77], DLBCL samples were classified as of germinal cell like (GCB) or non-GCB types.

Ethics

The study followed the tenets of the Helsinki Declaration and was approved by the local ethics committee (Journal no. H-B-2009-054) and the Danish Data Protection Agency (Journal no. 2009-41-3281). The authors declare no conflicts of interest

RNA extraction

Total RNA was isolated from four to eight 10 µm tissue sections using the RecoverAll, Total Nucleic Acid Isolation Kit (Applied Biosystems/Ambion, USA) according to manufacturer guidelines. Briefly, the samples were incubated in Xylene at 50°C to remove paraffin excess, followed by ethanol washes. Proteins were degraded by digestion buffer using added protease solution at 50/80°C. Samples were bound to a spin-filter by addition of isolation buffer and ethanol. DNA was degraded by DNAse treatment. The filter was washed several times and total RNA was eluted in 60 µl elution solution. Total RNA quantity and quality were checked by spectrophotometer (Nanodrop ND-1000).

MicroRNA microarray

From each sample 500 ng of total RNA was labeled with Hy3 fluorescent dye using the miRCURY LNA Array power labeling kit (Exiqon, Denmark). In brief, total RNA was mixed with a spike-in mix containing 52 different synthetic unlabeled miRNAs and flowingly treated with Calf Intestinal Alkaline Phosphatase (CIP) to catalyze the 5' phosphate group's removal in order to prevent self-ligation for 30 min at 37°C. The CIP treated RNA was labeled with Hy3 by addition of labeling buffer, fluorescent Hy3 label, DMSO and labeling enzyme, and incubated at 16°C. All samples were labeled the same day with the same master mix, in order to minimize technical variation.

The Hy3-labelled samples were hybridized to miRCURY LNA arrays (v11.0) (Exiqon, Denmark), which contains capture probes targeting all human miRNAs registered in the miRBASE version 15.0 at the Sanger Institute as well as a selection of other small RNAs. The hybridization was performed according to manufacturer specifications using a Tecan HS4800 hybridization station (Tecan, Austria). Slides are loaded into the Tecan Hybridization station and primed with 1 x hybridization buffer. 50 µl sample is injected and the inlet is flushed with 10µl 1x hybridization buffer. Sample is hybridized to the slide at 56°C for 16 hrs. at medium agitation. The slides are washed several times at room temperature with wash buffers of varying salt- and detergent concentration. Finally the slides are dried using pure nitrogen. Since it was not possible to hybridize all arrays in one go, samples were randomly spilt into 5 batches as to minimize day-to-day variation in the hybridization process.

After hybridization the microarray slides were scanned using the Agilent G2565BA Microarray Scanner System (Agilent Technologies, Inc., USA) at 5µm resolution and the resulting TIFF images segmented using the ImaGene 8.0 software on standard settings (BioDiscovery, Inc., USA).

Reverse transcription and Quantitative PCR (RT qPCR)

MicroRNAs were quantified using the miRCURY LNA™ Universal RT microRNA PCR kit (Exiqon, Hørsholm, Denmark) according to manufacturer's protocol. Briefly, reverse transcription (RT) was performed using 40 ng of total RNA from each sample. For quality control of the cDNA synthesis, RNA spike-in was added to each RT reaction, and later quantified using Control primers included in the kit. The miRNAs let-7g, miR-16, miR-24-2*, miR-27a, miR-27b, miR-29a, miR-29b, miR-29c, miR-30e, miR-122, miR-199a, miR-320b, miR-1248 and miR-1908 (nomenclature according to miRBase 18) were quantified using LNA^{TM} PCR primer sets, and the reference RNAs SNORD44 and SNORD48 were used for normalization of input RNA. RT-qPCR was performed using the LightCycler 480 instrument (Roche diagnostics, Basel, Switzerland) and the conditions recommended by

Exiqon for 42 lymphoma samples. The PCR cycle crossing point (C_p) was determined for each well and C_p estimates from two technical replicates for each probe-set were combined to produce one signal by taking the mean of reliable wells.

Analysis of miRNA expression array

All microarray data processing steps were performed using the R environment [143]. The Genepix median intensities were read into R using LIMMA package [172]. Further, the *normexp* method of background correction (with offset=10) was applied [152] to correct for spatial biases. This step was followed by within-array normalization using global loess. To reduce variance between arrays, we applied cubic spline normalization [152, 173]. For the differential expression analysis between eDLBCL and eMALT, only foreground intensities were used. Empirical Bayes moderated t-test statistic were obtained using LIMMA package [172, 170]. P-values were adjusted using the Benjamini-Hoeschberg correction method [7].

MicroRNA regulatory network

Predicted and validated set of targets of the significantly differentially expressed miRNA were obtained from MiRwalk [44]. Transcription factor (TF) - miRNA interactions were obtained from transmiR database [186].

Results

Cohort summary

Of the 19 patients with eMALT, 10 were men. The median age was 70 years (range 26-90 years). Thirteen patients (79%) presented with Stage I lymphoma and six patients (21%) had Stage IV lymphoma. Information of treatment methods was available in 12 patients. Of these, 9 (75%) patients were treated with radiation therapy, and 3 (25%) received systemic chemotherapy. All patients achieved complete remission and none had a relapse. The 5-year overall survival for the entire population was 79%.

The 25 patients with ocular DLBCL had male predominance (15/25, 60%), and the median age was 78 years (range 35-97 years). Thirteen patients (52%) presented with Stage I lymphoma. Of these 12 (92%) were diagnosed with primary Lymphoma. Three patients (12%) had Stage II, one patient (4%) had Stage III and 8 patients (32%) presented with Stage IV lymphoma. Nine of the DLBCL patients were classified as GCB and sixteen as non-GCB type. Systemic chemotherapy was applied in 16 of the patients (68%), seven patients (28%) were treated with radiation therapy, and one patient (4%) had surgery. Sixteen of the 25 patients experienced a relapse.

Table 7.1. Differentially expressed miRNAs for eDLBCLs relative to eMALT, log-fold-changes (LFC) and their associated p-values (adjusted p-value < 0.01)

miRNA	LFC	P-value	miRNA	LFC	P-value
Down-regulated miRNAs					
miR-24-2*	-2.45	6.98E-44	miR-222	-0.98	3.26E-19
miR-29a	-2.12	1.97E-31	miR-27a	-0.97	2.57E-17
miR-29b	-2.10	5.53E-28	miR-342-3p	-0.96	5.72E-15
miR-24-1*	-2.08	7.49E-44	miR-548g	-0.96	3.35E-05
miR-29c	-1.93	1.96E-26	miR-320c	-0.95	4.85E-20
miR-26a	-1.62	7.42E-27	miR-1297	-0.93	4.93E-24
miR-26b	-1.62	1.65E-29	miR-1200	-0.90	1.02E-09
let-7g	-1.56	6.50E-34	miR-137	-0.88	2.27E-05
miR-140-3p	-1.29	1.04E-29	miR-708	-0.87	4.72E-30
miR-142-3p	-1.27	7.48E-17	miR-23b	-0.87	1.95E-10
miR-451	-1.18	2.51E-08	miR-27b	-0.87	2.52E-13
miR-27a*	-1.15	7.42E-27	let-7a	-0.85	2.81E-13
miR-142-5p	-1.12	5.01E-13	miR-145	-0.84	6.34E-11
miR-16	-1.10	2.41E-18	miR-24	-0.83	1.86E-11
miR-125b	-1.09	3.07E-10	miR-30c	-0.82	5.82E-18
miR-30e	-1.07	2.64E-13	miR-1248	-0.81	4.43E-16
miR-101	-1.06	2.30E-16	miR-30b	-0.80	8.89E-12
miR-199a-5p	-1.01	5.51E-09	miR-221	-0.79	3.88E-15
miR-199a-3p	-1.01	4.66E-08	let-7f	-0.74	3.22E-11
miR-645	-1.00	7.41E-18	miR-98	-0.73	2.65E-10
miR-143	-0.99	1.31E-09			
Up-regulated miRNAs					
miR-663b	0.71	5.44E-06			
miR-30b*	0.85	6.66E-09			

The 5-year overall survival the entire group was 13%.

Comparison of miRNA expression profiles in MALT and DLBCL

Out of the 1900 RNA species analyzed on the array, 892 human mature-miRNAs (miRBase release 14.0) were investigated for 43 lymphoma samples. No clear correlation was observed between the known immuno-stratification of the samples, according to the HANS classification, and the miRNA expression profiles, as seen in Figure 7.1. Differential expression analysis identified 43 significantly differentially expressed miRNAs in eDLBCL vs. eMALT, 2 up-regulated and 41 down regulated in DLBCL relative to MALT, see Table 7.1.

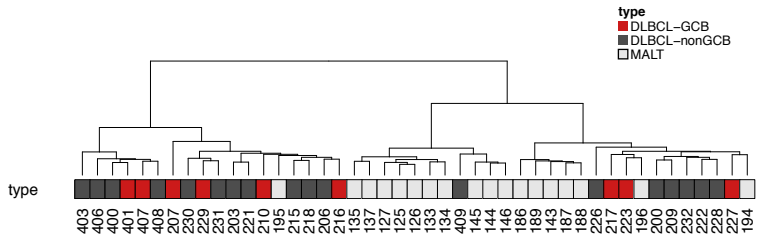

Figure 7.1. Two-dimensional hierarchical clustering of ocular MALT-type and diffuse large B-cell lymphomas shown for an un-biased selection of miRNA genes. The dendrogram represents a clustering of 892 miRNAs using ward linkage hierarchical clustering and the Euclidean distance between mean centered and variance scaled miRNA profiles, i.e. miRNA Z-scores.

Supervised hierarchical clustering of the significantly differential expressed miRNAs

To further characterize the miRNAs that were differentially expressed between eMALT and eDLBCLs we performed a supervised clustering. Using this approach only two outliers were observed (Figure 7.2). Although the analysis of histology did not reveal any characteristic features of these two tumors, sample *186* of eMALT type showed a higher level of the proliferation marker KI67. The other outlier (sample *228*) was a DLBCL that clustered with the MALT type lymphomas. Interestingly this sample had a relatively low proliferation marker of 30%.

RT-qPCR based validation was performed for 15 miRNAs, including 12 miRNAs that were found to be differentially expressed using microarrays (let-7g, miR-16, miR-29a, miR-29b, miR-29c, miR-199a-5p, miR-222, miR-24, miR-27a, miR-27b, miR-30e, miR-1248); miR-129-5p and miR-1908 which were not found to be differentially expressed; miR-320b, that differs in mature RNA sequence only slightly from miR-320c which was differentially expressed. The Pearson correlation coefficients (PCC) between microarray and qPCR measurements ranged from –0.66 to -0.95, due to inverse relationship between C_p values and miRNA concentrations for 11 of the 13 tested miRNAs. All 13 miRNA correlation coefficients in this range were found to be significant ($p < 0.01$), while miR-1908 and the negative control (miR-129-5p) were not found to be significant (CC0). Of the 13, 9 exhibited PCC < -0.90 ($p <$ 1e-15) and 2 exhibited $-0.86 < PCC < -0.66$ ($p <$ 1e-5). Correlations analysis and PCC values can be seen in Supplementary Figure 7.2. Of the 12 significantly differentially expressed miRNAs found using microarrays, qPCR confirmed differential expression of 10 (let-7g, miR-16,

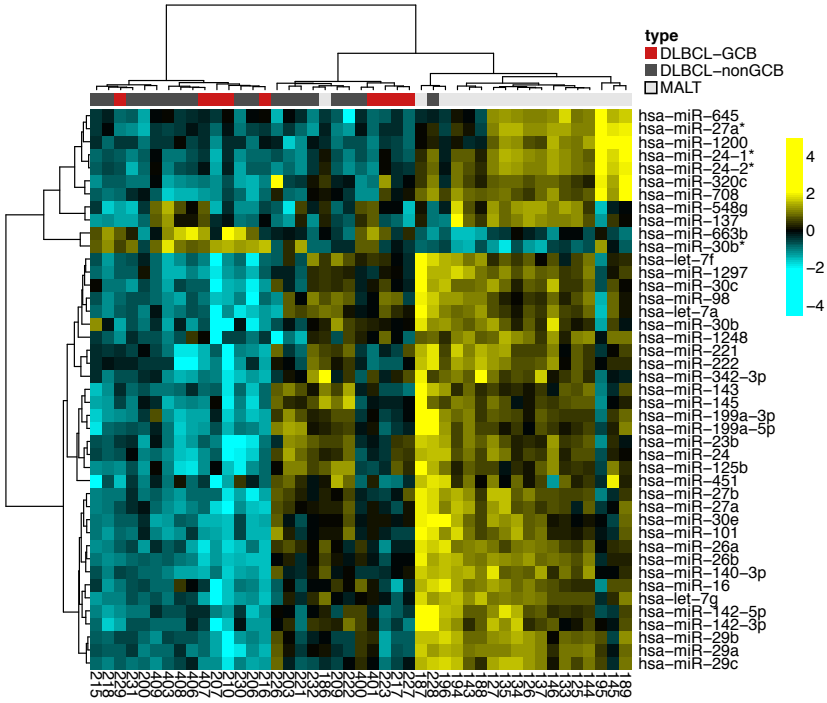

Figure 7.2. Two-dimensional hierarchical clustering of ocular MALT-type and diffuse large B-cell lymphomas for the 43 significantly differentially expressed miRNAs between MALT and DLBCL. The 41 miRNAs down-regulated in DLBCL show a less variable pattern across patient samples relative to the 2 miRNAs up-regulated in DLBCL

miR-29a, miR-29b, miR-29c, miR-199a-5p, miR-222, miR-27a, miR-30e and miR-1248) as shown in Figure 7.2. Points outside the grey area pass the significance thresholds for logFC (absolute $\log_2$FC$\geq$0.7 and p-value$\leq$0.05). The overall correlation in terms of fold-changes and the p-values were -0.636 and 0.75, respectively. In summary, an excellent correlation between array and qPCR results was observed.

Suppression of MYC regulated miRNAs in diffuse large B-cell lymphoma

MYC, a transcription factor (TF) and a lymphoma-associated oncoprotein, controls various steps of normal B-cell development. Dysregulation or a gain-of-function of MYC in various human malignancies, particularly in lymphomagenesis, suggests its crucial role in regulation of protein-coding

and non-coding genes in tumorigenesis. TransmiR [186], a database of TF-miRNA regulation, reports that MYC regulates 38 miRNAs (MYC-miR interactions), of which 24 miRNAs are known to negatively regulated, 10 miRNAs to be positively regulated and 4 are under an undetermined regulatory influence of MYC. the down-regulated miRNAs, let-7g, let-7f, let-7a, miR-23b, miR-26a, miR-29a and miR-29c have been reported to be under direct negative regulation of MYC [26, 27, 162, 166], as seen in Figure 7.4. Further, MYC has been observed to bind the promoter upstream of tumor suppressor let-7 family pri-miRNAs [26]. Target predictions obtained for let-7 family miRNAs and miR-26a include MYC and other cancer-related genes, such as BCL2, BCL6, BRAF, EGFR, DICER1, KRAS, MAP2K4, MYCN, NRAS, RUNX1, TNFAIP3, TNFRSF14 and TP53. This implies a negative feedback loop between MYC and the MYC-repressed miRNAs. Such negative feedback loops result in the loss of post-transcriptional regulation of let-7 family and miR-26a targets, leading to up-regulation of oncogenes. Targets of MYC-repressed miRNAs, together with MYC-induced oncogenes, may contribute to tumorigenesis and transformation of eMALT to eDLBCL. Our finding is in agreement with previous study [38] on transformation of gastric MALT to gastric DLBCL, which also suggested a role for MYC dysegulation in MALT-DLBCL transformation.

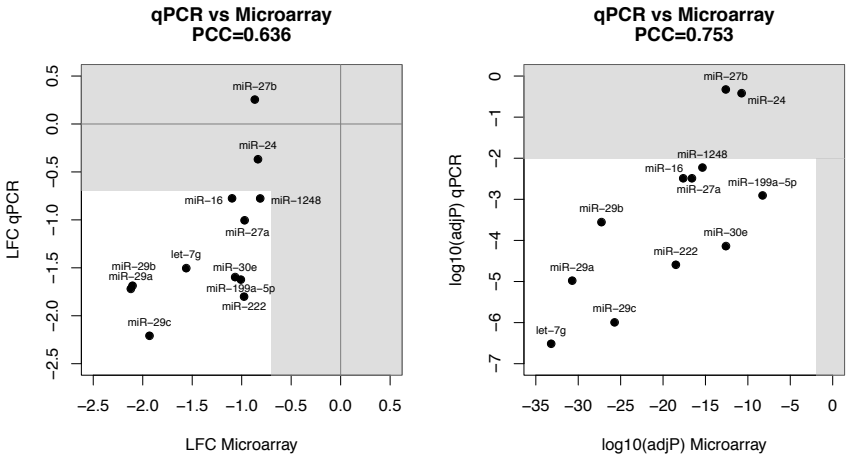

Figure 7.3. Agreement of the differential expression and corresponding p-values measured by microarrays and RT-qPCR. Points outside the gray rectangles represent miRNAs that pass the $\log_2$ fold-change and p-value thresholds (absolute $\log_2$ fold-change $\geq$0.7 and p-value$\leq$0.05

Activation of NF-KB signaling via loss of miRNA-regulation

Target prediction for the differentially expressed miRNAs revealed a small set of miRNAs regulating NFKB1. These miRNAs include, let-7g, miR-1275, miR-199a-5, miR-26a, miR-27a* and miR-342. All, except for miR-1275, showed down-regulation in eDLBCL relative to eMALT samples. This hints toward a loss of regulation of NFKB1 through its regulatory miRNAs. We suggest that an aberrant activation of NFKB signaling through either chromosomal aberrations, or due to a loss of miRNA-mediated regulation may lead to an over-expression of NFKB1. Furthermore, NFKB1 also transcriptionally regulates the expression of miR-29 family members and miR-125b NFKB1. It has been previously shown that NFKB1 suppresses transcription of miR-29a/b/c and miR-125b by binding to their upstream promoters [185]. In agreement with this plausible mechanism, we observed significant down-regulation of these miRNAs in eDLBCL relative to eMALT (log2 fold-changes: miR-29a/b/c < -1.9 and -1.08 for miR-125b), see Figure 7.4.

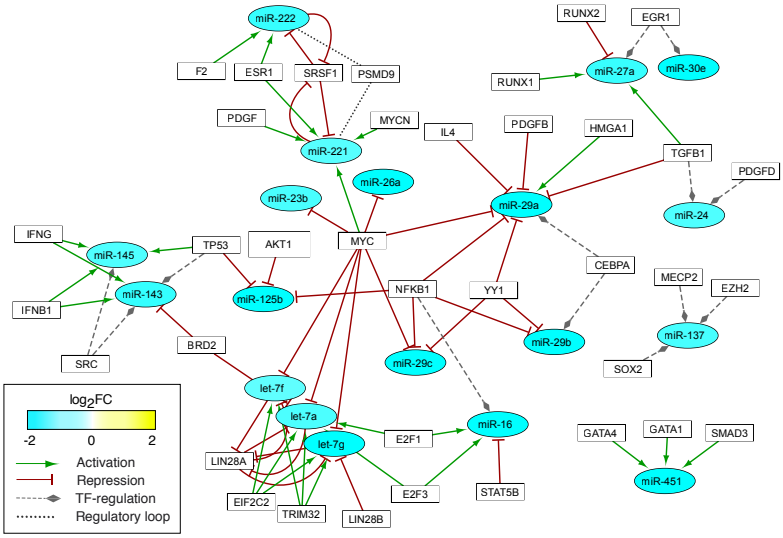

Figure 7.4. TF-miRNA interactions for the most significantly differentially expressed miRNAs in eDLBCL as curated by transmiR database

Discussion

Ocular adnexal lymphoma is a group of malignancies with heterogeneous clinical, molecular and pathological characteristics. MALT in ocular regions

is one of the most frequent types of lymphomas in eye that despite being indolent initially, often transform into the malignant high-grade diffuse large B-cell lymphoma of eye. In the present study we investigated the role of miRNAs in disease transformation from eMALT to eDLBCL by miRNA expression profiling of FFPE specimens from patients suffering from these diseases.

We observed 43 differentially expressed miRNAs between eMALT and eDLBCL samples, a majority of which (41) were found at lower levels in DLBCL than in MALT. A selected set of these miRNAs was validated by RT-qPCR on all patient material. Of the thirteen miRNAs for which qPCR validations were performed, 9 showed a significant correlation to their expression profiles (PCC < -0.90, p < 1e-15) and 2 exhibited a PCC ranging from -0.86 to -0.66 (p < 1e-5). miR-29 family miRNAs, miR-222, let-7g and miR-199-5p showed highest and most significant correlations between their gene expression signatures and C_p values. The differential expression of the selected miRNAs measured by microarrays was in agreement with the one measured using qPCR (correlation of $\log_2$ fold-change, -0.636 and of the p-values, 0.75).

The miRNAs significantly down-regulated in DLBCL included miR-29a, miR-29a* and miR-24 family of miRNAs; and miR-221/222, miR-23a and miR-29 cluster. miR-29a/c have been previously reported to be under a negative transcriptional regulation by MYC and NFKB1. We also observed a significant down-regulation of other known MYC suppressed miRNAs such as miR-26a, let-7g, and miR-221. let-7g is also repressed by LIN28B, a protein that is trans-activated by MYC. However, it has been shown that the presence of Lin28B is a necessary and sufficient condition for MYC-mediated repression of let-7g and does not change the expression of pri-let-7 [26]. In summary, at least 7 of 29 down-regulated miRNAs are known to be under transcriptional repression by MYC. Interestingly, known upstream regulators of MYC, miR-142-3p/5p, were one of the most down-regulated miRNAs (more than two-fold down-regulation). Thus, a loss-of-function of miR-142 may have resulted in MYC over expression, and thereby a repression of several tumor suppressive miRNAs.

Aberrant activation of NF-κB signaling pathway has repeatedly been linked to the aggressive forms of diffuse large B-cell lymphoma [110]. Such aberrant activation in MALT and DLBCL may result from recurrent chromosomal anomalies, such API2/MALT1 and IGH/MALT1 translocations, as previously reported in 30% to 50% of extranodal MALT lymphomas [81]. The oncogenic fusion protein API2/MALT1 or an IGH-mediated activation of MALT1 due to IGH/MALT1, results in to hyperactivation of NF-Kb signaling pathway. However, genetic regulatory mechanisms upstream of NFKB1, a key gene for NFKB signaling event, may also lead to an aberrant expression and activation of NF-KB pathway. We observed that four out of

five regulatory miRNA of NFKB1 showed a significant down regulation in eDLBCL, suggesting a loss-of-regulation event through upstream miRNA-regulators of NFKB1. Therefore, we hypothesize that NFKB1 expression might exhibit an increase after its escape from miRNA-mediated regulation. Interestingly, we also observe that NFKB1 regulated miRNAs, miR-29a/b/c were also significantly down-regulated in DLBCL. NFKB1 has been reported as a transcriptional regulatory of miR-29a/b/c and it has negative regulatory effect. In addition, the largest down-regulation events were observed for the miR-29 family, which is also transcriptionally suppressed by YY1 (ying-yang 1), a zinc finger protein, over-expression of which has been linked with a shorter survival in DLBCL and follicular lymphomas [185, 161]. This intriguing finding deserves experimental validation of mRNA or protein levels of MYC, NFKB1 and YY1 in freshly frozen samples as these measurements are, unfortunately, not possible with FFPE samples.

Among the down-regulated miRNAs, miR-16, miR-221, miR-26b, miR-24, miR-23b, miR-27b, and miR-30e have previously been linked with B-cell lymphomas [26, 197] and other lymphoproliferative disorders [20]. Further, the down-regulated miRNAs were enriched for tumor suppressive miRNAs, let-7a/f/g, miR-16, miR-451, miR-26a/b, miR-145, miR-143, miR-125b, miR-101, miR-29a/c and miR-98. This finding suggests a systematic repression of tumor suppressor miRNAs during the transformation of MALT to DLBCL.

In summary, a large scale repression of tumor suppressive miRNAs observed during the transformation of eMALT to eDLBCL provides hints toward their potential in diagnostics and therapeutics.

Supporting Information

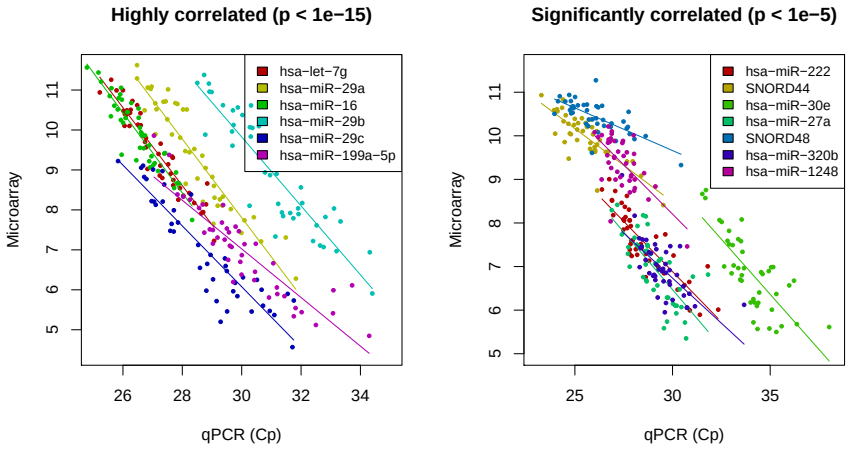

Figure 7.5. Correlation plots showing good correlation between microarray and RT-qPCR determined miRNA levels

Part III

Epilogue

Chapter 8

Summary & future perspectives

This thesis presents projects that elucidate the role of miRNAs in drug resistance in ER-positive breast cancers and in the disease transformation from ocular MALT to ocular DLBCL. The projects herein, describe analyses of data obtained using various gene expression profiling technologies. The statistical methods are applied such that to reduce the technical variation as much as possible and to highlight the biological differences between sample types. Further, the projects describe herein, integrate various lines of evidence coming from gene regulation by miRNAs, TF-miRNA, TF-DNA and protein-protein interactions.

Patients with ER-positive breast cancer are often given an adjuvant treatment with antiestrogen drugs such as Tamoxifen.Tamoxifen is highly effective in reducing the risk of recurrence after primary treatment, as well as in reducing risk of developing the disease for the patients with high-risk of developing breast cancer. Chapter 6 presents an integrative analysis of miRNA-target relationships for tamoxifen resistance in ER-positive breast cancer. Inverse correlations between miRNA and their targets were exploited to derive functional miRNA-target relationships. The miRNA-regulated genes were investigated for the pathways they might influence. From the functional enrichment analysis of miRNA-regulated differentially expressed genes, we determined several receptor signaling events, aurora B kinase signaling, FOXM1 signaling, etc. were affected by miRNA-mediated regulation. Further, with the integration of existing knowledge on TF-mediated regulation of miRNAs and protein coding genes, a linear pathway involving FOXM1, miR-135a and EPHA4/7 was identified. In addition, this study also highlights several detailed lines of events possibly contributing to the tamoxifen resistance.

Ocular adnexal lymphomas are the most common type of malignancies in the eye, covering intraocular and orbital (eyelid, lacrimal glands, conjunctiva, etc) regions. Mucosa associated lymphoid tissue lymphoma (MALT) and diffuse large B-cell lymphoma (DLBCL) is the most common extranodal malignancies in the eye. When untreated, MALT, a relatively indolent disease often transforms into the malignant DLBCL. Chapter 7 presents a project in which global alterations in the miRNA expression during the transformation of ocular MALT to ocular DLBCL were studied. Majority of these alterations referred to the repressed levels of tumor suppressor miRNAs in DLBCL. Dysregulated miR-29a/c, let-7a, let-7g, miR-23b, miR-26a and miR-221 are known to be under MYC-mediated transcriptional repression. miR-29 family members, down-regulated in DLBCL, are also known to be the transcriptional targets of NFKB1 and YY1. Differential expression of 10 out of 12 differentially expressed miRNAs was validated and confirmed by qPCR. Several other repressed miRNAs had a tumor suppressive effect. The study indicates a potential use of the dysregulated miRNAs in diagnostics or in therapeutics.

Experimental design is the most crucial step if the statistically and biologically valid conclusions are to be drawn from the data. The project described in Chapter 6 began with a miRNA expression profiling using small RNA-seq technologies. Various read alignment strategies were attempted to highlight the biological variation between cell lines as much as possible, differential expression analysis was carried out. However, in the absence of replicate measurements on the same cell lines, tests for differential expression did not yield statistical power and were insufficient to draw a sound conclusion. Therefore, a miRNA qPCR profiling with three replicate measurements were performed. In order to understand the impact of miRNA expression, we integrated microarray data obtained on the same RNA. The inverse correlation analysis, as outlined in the manuscript I, should ideally be performed with as many measurements available for miRNA and gene expression measurements. However, since the gene expression profiling experiment was planned prior to miRNA qPCR, we did not have matching number of samples for each cell line. Therefore, the mean expressions of replicate measurements for each cell line were obtained to perform an inverse correlation analysis. This may have dampened positive correlations between miRNA and mRNA pairs, and hence the analysis may have missed functional miRNA-target pairs. For such type of analysis of inferring miRNA-target relationships, we would like to plan the experiments such that we have the matching number of replicates for each sample type for miRNA and mRNA datasets.

For the project described in Chapter 7, the primary goal was to identify miRNAs with altered expression between ocular MALT and ocular DLBCL. Since DLBCL is a heterogeneous disease, the secondary goal was to obtain miRNA signatures that may identify molecular sub-groups of DLBCL. We applied an adaptive gene shaving approach to DLBCL dataset consisting of

25 patient samples (described in Chapter 5). We were able to successfully identify sub-groups of DLBCL samples with distinct miRNA expression signatures. However, the correlation of the sample sub-groups to clinical characteristics did not yield statistically significant associations. We account the smaller population (N=25) and shorter patient survivals (median age $\geq$ 75 years) for the lack of statistical significance. Although not the goal of the study, it would be very interesting to study gene expression changes and chromosomal aberrations together with changes in miRNA expression during the transformation from MALT to DLBCL. Unfortunately, the available patient samples are FFPE (formalin fixed paraffin embedded), and therefore it is rather difficult to obtain good quality mRNA from them.

In summary, this thesis presents studies that point toward various crucial roles of miRNAs in anti-estrogen resistance and in disease transformation. The results obtained by following the systems biology approach provide hints toward several biological events that may carry therapeutic and diagnostic potential.

Bibliography

[1] Affymetrix (2010). Design and Performance of the GeneChip® Human Genome U133 Plus 2.0 and Human Genome U133A 2.0 Arrays. Technote. 25

[2] Affymetrix (2010). GeneChip Expression Analysis Technical Manual. 25

[3] Ambros V., Bartel B., Bartel D.P., Burge C.B., Carrington J.C., et al. (2003). A uniform system for microRNA annotation. *RNA*, 9(3):277–279. 5

[4] Anders S. and Huber W. (2010). Differential expression analysis for sequence count data. *Genome Biology*, 11(10):R106. 43

[5] Arora A. and Simpson D.A. (2008). Individual mRNA expression profiles reveal the effects of specific microRNAs. *Genome Biology*, 9(5):R82. 8

[6] Baylin S.B. and Jones P.A. (2011). A decade of exploring the cancer epigenome - biological and translational implications. *Nature Reviews. Cancer*, 11(10):726–734. 75

[7] Benjamini Y. (2001). The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics*, 29(4):1165–1188. Mathematical Reviews number (MathSciNet): MR1869245; Zentralblatt MATH identifier: 01829051. 39, 57, 78

[8] Bergamaschi A., Christensen B.L., and Katzenellenbogen B.S. (2011). Reversal of endocrine resistance in breast cancer: interrelationships among 14-3-3 , FOXM1, and a gene signature associated with mitosis. *Breast Cancer Research*, 13(3):R70–R70. 64

[9] Bergamaschi A. and Katzenellenbogen B.S. (2012). Tamoxifen downregulation of miR-451 increases 14-3-3 and promotes breast cancer cell survival and endocrine resistance. *Oncogene*, 31(1):39–47. 64

[10] Betel D., Koppal A., Agius P., Sander C., and Leslie C. (2010). Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. *Genome Biology*, 11(8):R90. 8

[11] Betel D., Wilson M., Gabow A., Marks D.S., and Sander C. (2007). The microRNA.org resource: targets and expression. *Nucleic Acids Research*, 36(Database):D149–D153. 7

[12] Blow M.J., Grocock R.J., van Dongen S., Enright A.J., Dicks E., et al. (2006). RNA editing of human microRNAs. *Genome Biology*, 7(4):R27. 42

[13] Boguski M.S., Tolstoshev C.M., and Bassett J. D E (1994). Gene discovery in dbEST. *Science (New York, N.Y.)*, 265(5181):1993–1994. PMID: 8091218. 28

[14] Bolstad B.M., Irizarry R.A., Åstrand M., and Speed T.P. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*, 19(2):185–193. 33

[15] Booth C.S., Pienaar E., Termaat J.R., Whitney S.E., Louw T.M., et al. (2010). Efficiency of the polymerase chain reaction. *Chemical Engineering Science*, 65(17):4996–5006. 27

[16] Brantley-Sieders D.M., Jiang A., Sarma K., Badu-Nkansah A., Walter D.L., et al. (2011). Eph/Ephrin Profiling in Human Breast Cancer Reveals Significant Associations between Expression Level and Clinical Outcome. *PLoS ONE*, 6(9):e24426–e24426. 64

[17] Briand P. and Lykkesfeldt A.E. (1984). Effect of estrogen and antiestrogen on the human breast cancer cell line MCF-7 adapted to growth at low serum concentration. *Cancer Research*, 44(3):1114–1119. 54

[18] Bullard J.H., Purdom E., Hansen K.D., and Dudoit S. (2010). Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics*, 11(1):94. 43

[19] Burns N., Grimwade B., Ross-Macdonald P.B., Choi E.Y., Finberg K., et al. (1994). Large-scale analysis of gene expression, protein localization, and gene disruption in Saccharomyces cerevisiae. *Genes & Development*, 8(9):1087–1105. 23

[20] Calin G.A., Ferracin M., Cimmino A., Di Leva G., Shimizu M., et al. (2005). A MicroRNA signature associated with prognosis and progression in chronic lymphocytic leukemia. *The New England Journal of Medicine*, 353(17):1793–1801. PMID: 16251535. 85

[21] Calin G.A., Sevignani C., Dumitru C.D., Hyslop T., Noch E., et al. (2004). Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. *Proceedings of the National Academy of Sciences of the United States of America*, 101(9):2999–3004. PMID: 14973191. 8, 9

[22] Campo E., Swerdlow S.H., Harris N.L., Pileri S., Stein H., et al. (2011). The 2008 WHO classification of lymphoid neoplasms and beyond: evolving concepts and practical applications. *Blood*, 117(19):5019–32. Campo, Elias Swerdlow, Steven H Harris, Nancy L Pileri, Stefano Stein, Harald Jaffe, Elaine S Blood. 2011 May 12;117(19):5019-32. Epub 2011 Feb 7. 76

[23] Castellano L., Giamas G., Jacob J., Coombes R.C., Lucchesi W., et al. (2009). The estrogen receptor-Œ±-induced microRNA signature regulates itself and its transcriptional response. *Proceedings of the National Academy of Sciences*, 106(37):15732–15737. 53, 59

[24] CGHFBC C.G.o.H.F.i.B.C. (1996). Breast cancer and hormonal contraceptives: collaborative reanalysis of individual data on 53 297 women with breast cancer and 100 239 women without breast cancer from 54 epidemiological studies. Collaborative Group on Hormonal Factors in Breast Cancer. *Lancet*, 347(9017):1713–1727. PMID: 8656904. 12

[25] Chandra V., Girijadevi R., Nair A.S., Pillai S.S., and Pillai R.M. (2010). MTar: a computational microRNA target prediction architecture for human transcriptome. *BMC Bioinformatics*, 11(Suppl 1):S2. 8

[26] Chang T.C., Yu D., Lee Y.S., Wentzel E.A., Arking D.E., et al. (2008). Widespread microRNA repression by Myc contributes to tumorigenesis. *Nat Genet*, 40(1):43–50. Chang, Tsung-Cheng Yu, Duonan Lee, Yun-Sil Wentzel, Erik A Arking, Dan E West, Kristin M Dang, Chi V Thomas-Tikhonenko, Andrei Mendell, Joshua T R01 CA102709-01/CA/NCI NIH HHS/ R01 CA102709-02/CA/NCI NIH HHS/ R01 CA102709-03/CA/NCI NIH HHS/ R01 CA102709-04/CA/NCI NIH HHS/ R01 CA122334-01A1/CA/NCI NIH HHS/ R01 CA122334-02/CA/NCI NIH HHS/ Nat Genet. 2008 Jan;40(1):43-50. Epub 2007 Dec 9. 82, 84, 85

[27] Chang T.C., Zeitels L.R., Hwang H.W., Chivukula R.R., Wentzel E.A., et al. (2009). Lin-28B transactivation is necessary for Myc-mediated let-7 repression and proliferation. *Proc Natl Acad Sci U S A*, 106(9):3384–9. Chang, Tsung-Cheng Zeitels, Lauren R Hwang, Hun-Way Chivukula, Raghu R Wentzel, Erik A Dews, Michael Jung, Jason Gao, Ping Dang, Chi V Beer, Michael A Thomas-Tikhonenko, Andrei Mendell, Joshua T R01CA102709/CA/NCI NIH HHS/ R01CA120185/CA/NCI NIH HHS/ R01CA122334/CA/NCI NIH HHS/ Proc Natl Acad Sci U S A. 2009 Mar 3;106(9):3384-9. Epub 2009 Feb 11. 82

[28] Chanudet E., Zhou Y., Bacon C.M., Wotherspoon A.C., M√ºller,ÄêHermelink H., et al. (2006). Chlamydia psittaci is variably associated with ocular adnexal MALT lymphoma in different geographical regions. *The Journal of Pathology*, 209(3):344–351. 21

[29] Chen Y., Dougherty E.R., and Bittner M.L. (1997). Ratio-Based Decisions and the Quantitative Analysis of cDNA Microarray Images. 38

[30] Cheng C. and Li L.M. (2008). Inferring MicroRNA Activities by Combining Gene Expression with MicroRNA Target Prediction. *PLoS ONE*, 3(4):e1989. 8

[31] Chien C.H., Sun Y.M., Chang W.C., Chiang-Hsieh P.Y., Lee T.Y., et al. (2011). Identifying transcriptional start sites of human microRNAs based on high-throughput sequencing data. *Nucleic Acids Research*. 4

[32] Cittelly D.M., Das P.M., Spoelstra N.S., Edgerton S.M., Richer J.K., et al. (2010). Downregulation of miR-342 is associated with tamoxifen resistant breast tumors. *Molecular Cancer*, 9(1):317–317. 65

[33] Cock P.J.A., Fields C.J., Goto N., Heuer M.L., and Rice P.M. (2010). The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Research*, 38(6):1767–1771. 39

[34] Coombes K.R., Tsavachidis S., Morris J.S., Baggerly K.A., Hung M., et al. (2005). Improved peak detection and quantification of mass spectrometry data acquired from surface-enhanced laser desorption and ionization by denoising spectra with the undecimated discrete wavelet transform. *Proteomics*, 5(16):4107–4117. PMID: 16254928. 35

[35] Corney D.C., Flesken-Nikitin A., Godwin A.K., Wang W., and Nikitin A.Y. (2007). MicroRNA-34b and MicroRNA-34c Are Targets of p53 and Cooperate in Control of Cell Proliferation and Adhesion-Independent Growth. *Cancer Research*, 67(18):8433–8438. 8

[36] Couch F.J., Gaudet M.M., Antoniou A.C., Ramus S.J., Kuchenbaecker K.B., et al. (2012). Common Variants at the 19p13.1 and ZNF365 Loci Are Associated with ER Subtypes of Breast Cancer and Ovarian Cancer Risk in BRCA1 and BRCA2 Mutation Carriers. *Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology*. PMID: 22351618. 11

[37] Cowland J.B., Hother C., and Gr√ nbaek K. (2007). MicroRNAs and cancer. *APMIS: Acta Pathologica, Microbiologica, Et Immunologica Scandinavica*, 115(10):1090–1106. 75

[38] Craig V.J., Cogliatti S.B., Imig J., Renner C., Neuenschwander S., et al. (2011). Myc-mediated repression of microRNA-34a promotes high-grade transformation of B-cell lymphoma by dysregulation of FoxP1. *Blood*, 117(23):6227–6236. 8, 73, 75, 82

[39] Daibata M., Nemoto Y., Togitani K., Fukushima A., Ueno H., et al. (2006). Absence of Chlamydia psittaci in ocular adnexal lymphoma from Japanese patients. *British Journal of Haematology*, 132(5):651–652. 21

[40] Danilovich N., Babu P.S., Xing W., Gerdes M., Krishnamurthy H., et al. (2000). Estrogen Deficiency, Obesity, and Skeletal Abnormalities in Follicle-Stimulating Hormone Receptor Knockout (FORKO) Female Mice. *Endocrinology*, 141(11):4295–4308. 14

[41] Dean J.D., Goodwin P.H., and Hsiang T. (2002). Comparison of relative RT-PCR and northern blot analyses to measure expression of -1,3-glucanase inNicotiana benthamiana infected withColltotrichum destructivum. *Plant Molecular Biology Reporter*, 20(4):347–356. 26

[42] Dudoit R., Yang Y.H., Callow M.J., and Speed T.P. (2002). Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments. *STATISTICA SINICA*, 12:111—139. 38

[43] Dweep H., Sticht C., Pandey P., and Gretz N. (2011). miRWalk–database: prediction of possible miRNA binding sites by "walking" the genes of three genomes. *Journal of Biomedical Informatics*, 44(5):839–847. 8, 57

[44] Dweep H., Sticht C., Pandey P., and Gretz N. (2011). miRWalk–database: prediction of possible miRNA binding sites by "walking" the genes of three genomes. *Journal of Biomedical Informatics*, 44(5):839–847. 78

[45] (EBCTCG) E.B.C.T.C.G. (2005). Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. *The Lancet*, 365(9472):1687–1717. 53

[46] Ebert M.S., Neilson J.R., and Sharp P.A. (2007). MicroRNA sponges: competitive inhibitors of small RNAs in mammalian cells. *Nature Methods*, 4(9):721–726. 9

[47] El Baroudi M., Corà D., Bosia C., Osella M., and Caselle M. (2011). A Curated Database of miRNA Mediated Feed-Forward Loops Involving MYC as Master Regulator. *PLoS ONE*, 6(3):e14742. 8

[48] Elias D., Hansen H., Laenkholm A.V., Yde C.W., Lykkesfeldt A., et al. (2012). Fyn play a key role in tamoxifen resistance in breast cancer and predict response to adjuvant tamoxifen treatment in patients. Manuscript in preparation. 55

[49] Esquela-Kerscher A. and Slack F.J. (2006). Oncomirs [mdash] microRNAs with a role in cancer. *Nat Rev Cancer*, 6(4):259–269. 75

[50] Ewing B., Hillier L., Wendl M.C., and Green P. (1998). Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Research*, 8(3):175–185. PMID: 9521921. 39

[51] Fahlgren N., Howell M.D., Kasschau K.D., Chapman E.J., Sullivan C.M., et al. (2007). High-Throughput Sequencing of <italic>Arabidopsis</italic> microRNAs: Evidence for Frequent Birth and Death of <italic>MIRNA</italic> Genes. *PLoS ONE*, 2(2):e219. 43

[52] Fare T.L., Coffey E.M., Dai H., He Y.D., Kessler D.A., et al. (2003). Effects of Atmospheric Ozone on Microarray Data Quality. *Analytical Chemistry*, 75(17):4672–4675. 24

[53] Fernandez S. and Russo J. (2010). Estrogen and Xenoestrogens in Breast Cancer. *Toxicologic pathology*, 38(1):110–122. PMID: 19933552 PMCID: 2907875. 12

[54] Ferreri A.J.M., Guidoboni M., Ponzoni M., De Conciliis C., Dell'Oro S., et al. (2004). Evidence for an association between Chlamydia psittaci and ocular adnexal lymphomas. *Journal of the National Cancer Institute*, 96(8):586–594. 21

[55] Ferry J.A., Fung C.Y., Zukerberg L., Lucarelli M.J., Hasserjian R.P., et al. (2007). Lymphoma of the Ocular Adnexa: A Study of 353 Cases. *The American Journal of Surgical Pathology*, 31(2):170–184. 21

[56] Feuer E.J., Wun L.M., Boring C.C., Flanders W.D., Timmel M.J., et al. (1993). The lifetime risk of developing breast cancer. *Journal of the National Cancer Institute*, 85(11):892–897. PMID: 8492317. 12

[57] Fisher B., Costantino J.P., Wickerham D.L., Redmond C.K., Kavanah M., et al. (1998). Tamoxifen for prevention of breast cancer: report of the National Surgical Adjuvant Breast and Bowel Project P-1 Study. *Journal of the National Cancer Institute*, 90(18):1371–1388. PMID: 9747868. 16

[58] Francis R.E., Myatt S.S., Krol J., Hartman J., Peck B., et al. (2009). FoxM1 is a downstream target and marker of HER2 overexpression in breast cancer. *International journal of oncology*, 35(1):57–68. 64

[59] Frasor J., Chang E.C., Komm B., Lin C.Y., Vega V.B., et al. (2006). Gene expression preferentially regulated by tamoxifen in breast cancer cells and correlations with clinical outcome. *Cancer Res*, 66(14):7334–40. Frasor, Jonna Chang, Edmund C Komm, Barry Lin, Chin-Yo Vega, Vinsensius B Liu, Edison T Miller, Lance D Smeds, Johanna Bergh, Jonas Katzenellenbogen, Benita S CA 18119/CA/NCI NIH HHS/United States Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't United States Cancer research Cancer Res. 2006 Jul 15;66(14):7334-40. 64

[60] Freeman C., Berg J.W., and Cutler S.J. (1972). Occurrence and prognosis of extranodal lymphomas. *Cancer*, 29(1):252–260. 19, 20, 75

[61] Friedman R.C., Farh K.K., Burge C.B., and Bartel D.P. (2009). Most mammalian mRNAs are conserved targets of microRNAs. *Genome Research*, 19(1):92–105. 7

[62] Gautier L., Cope L., Bolstad B.M., and Irizarry R.A. (2004). affy–analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics (Oxford, England)*, 20(3):307–315. 56

[63] Gerhard D.S., Wagner L., Feingold E.A., Shenmen C.M., Grouse L.H., et al. (2004). The status, quality, and expansion of the NIH full-length cDNA project: the Mammalian Gene Collection (MGC). *Genome Research*, 14(10B):2121–2127. PMID: 15489334. 28

[64] German M.A., Pillay M., Jeong D., Hetawal A., Luo S., et al. (2008). Global identification of microRNA-target RNA pairs by parallel analysis of RNA ends. *Nature Biotechnology*, 26(8):941–946. PMID: 18542052. 7

[65] Giacinti L., Claudio P.P., Lopez M., and Giordano A. (2006). Epigenetic Information and Estrogen Receptor Alpha Expression in Breast Cancer. *The Oncologist*, 11(1):1–8. 53, 63

[66] Git A., Dvinge H., Salmon-Divon M., Osborne M., Kutter C., et al. (2010). Systematic comparison of microarray profiling, real-time PCR, and next-generation sequencing technologies for measuring differential microRNA expression. *RNA*, 16(5):991–1006. 26

[67] Gökmen-Polar Y., Toroni R.A., Hocevar B.A., Badve S., Zhao Q., et al. (2010). Dual targeting of EphA2 and ER restores tamoxifen sensitivity in ER/EphA2-positive breast cancer. *Breast Cancer Research and Treatment*, 127(2):375–384. 61

[68] Griffiths-Jones S. (2006). miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Research*, 34(90001):D140–D144. 24, 42

[69] Griffiths-Jones S., Saini H.K., van Dongen S., and Enright A.J. (2008). miRBase: tools for microRNA genomics. *Nucleic Acids Research*, 36(Database issue):D154–158. PMID: 17991681. 7

[70] Grillari J., Hackl M., and Grillari-Voglauer R. (2010). miR-17,92 cluster: ups and downs in cancer and aging. *Biogerontology*, 11(4):501–506. 53

[71] Grimson A., Farh K.K., Johnston W.K., Garrett-Engele P., Lim L.P., et al. (2007). MicroRNA Targeting Specificity in Mammals: Determinants beyond Seed Pairing. *Molecular Cell*, 27(1):91–105. 6

[72] Grishok A., Pasquinelli A.E., Conte D., Li N., Parrish S., et al. (2001). Genes and Mechanisms Related to RNA Interference Regulate Expression of the Small Temporal RNAs that Control C. elegans Developmental Timing. *Cell*, 106(1):23–34. 3

[73] Grün D., Wang Y.L., Langenberger D., Gunsalus K.C., and Rajewsky N. (2005). microRNA Target Predictions across Seven Drosophila Species and Comparison to Mammalian Targets. *PLoS Comput Biol*, 1(1):e13–e13. 57

[74] Gu S. and Kay M.A. (2010). How do miRNAs mediate translational repression? *Silence*, 1(1):11. PMID: 20459656. 6

[75] Hamano R., Miyata H., Yamasaki M., Kurokawa Y., Hara J., et al. (2011). Overexpression of miR-200c induces chemoresistance in esophageal cancers mediated through activation of the Akt signaling pathway. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, 17(9):3029–3038. 53

[76] Hammell M., Long D., Zhang L., Lee A., Carmack C.S., et al. (2008). mirWIP: microRNA target prediction based on microRNA-containing ribonucleoprotein|[ndash]|enriched transcripts. *Nature Methods*, 5(9):813–819. 7, 8

[77] Hans C.P., Weisenburger D.D., Greiner T.C., Gascoyne R.D., Delabie J., et al. (2004). Confirmation of the molecular classification of diffuse large B-cell lymphoma by immunohistochemistry using a tissue microarray. *Blood*, 103(1):275–282. 76

[78] Hastie T., Tibshirani R., Eisen M.B., Alizadeh A., Levy R., et al. (2000). 'Gene shaving' as a method for identifying distinct sets of genes with similar expression patterns. *Genome Biology*, 1(2):research0003. 46

[79] Hayashita Y., Osada H., Tatematsu Y., Yamada H., Yanagisawa K., et al. (2005). A Polycistronic MicroRNA Cluster, miR-17-92, Is Overexpressed in Human Lung Cancers and Enhances Cell Proliferation. *Cancer Research*, 65(21):9628–9632. 8

[80] He L., Thomson J.M., Hemann M.T., Hernando-Monge E., Mu D., et al. (2005). A microRNA polycistron as a potential human oncogene. *Nature*, 435(7043):828–833. PMID: 15944707. 9

[81] Ho L., Davis R.E., Conne B., Chappuis R., Berczy M., et al. (2005). MALT1 and the API2-MALT1 fusion act between CD40 and IKK and confer NF-Œ B-dependent proliferative advantage and resistance against FAS-induced cell death in B cells. *Blood*, 105(7):2891–2899. 84

[82] Hoskins J.M., Carey L.A., and McLeod H.L. (2009). CYP2D6 and tamoxifen: DNA matters in breast cancer. *Nature Reviews. Cancer*, 9(8):576–586. PMID: 19629072. 16

[83] Howell S.J., Johnston S.R., and Howell A. (2004). The use of selective estrogen receptor modulators and selective estrogen receptor down-regulators in breast cancer. *Best Practice & Research Clinical Endocrinology & Metabolism*, 18(1):47–66. 14, 15

[84] Hua Z., Lv Q., Ye W., Wong C.A., Cai G., et al. (2006). MiRNA-Directed Regulation of VEGF and Other Angiogenic Factors under Hypoxia. *PLoS ONE*, 1(1):e116. 63

[85] Huse J.T., Brennan C., Hambardzumyan D., Wee B., Pena J., et al. (2009). The PTEN-regulating microRNA miR-26a is amplified in high-grade glioma and facilitates gliomagenesis in vivo. *Genes & Development*, 23(11):1327–1337. 8

[86] Illumina (2010). *Small RNA v1.5 Sample Preparation Guide*. Illumina Inc. 29

[87] Irizarry R.A., Hobbs B., Collin F., Beazer‚ÄêBarclay Y.D., Antonellis K.J., et al. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*, 4(2):249–264. 33, 34, 56

[88] Jaffe E.S. (2009). The 2008 WHO classification of lymphomas: implications for clinical practice and translational research. *Hematology / the Education Program of the American Society of Hematology. American Society of Hematology. Education Program*, pages 523–531. PMID: 20008237. 20

[89] John B., Enright A.J., Aravin A., Tuschl T., Sander C., et al. (2004). Human MicroRNA Targets. *PLoS Biol*, 2(11):e363–e363. 57

[90] Johnston S.R., Saccani-Jotti G., Smith I.E., Salter J., Newby J., et al. (1995). Changes in estrogen receptor, progesterone receptor, and pS2 expression in tamoxifen-resistant human breast cancer. *Cancer Research*, 55(15):3331–3338. 9, 53

[91] Ju J. (2010). miRNAs as biomarkers in colorectal cancer diagnosis and prognosis. *Bioanalysis*, 2(5):901–906. PMID: 20543966 PMCID: 2883182. 9

[92] Kai Z.S. and Pasquinelli A.E. (2010). MicroRNA assassins: factors that regulate the disappearance of miRNAs. *Nature Structural & Molecular Biology*, 17(1):5–10. 4

[93] Kent O.A., Chivukula R.R., Mullendore M., Wentzel E.A., Feldmann G., et al. (2010). Repression of the miR-143/145 cluster by oncogenic Ras initiates a tumor-promoting feed-forward pathway. *Genes & Development*, 24(24):2754–2759. 8

[94] Kent W.J. (2002). BLAT–the BLAST-like alignment tool. *Genome Research*, 12(4):656–664. PMID: 11932250. 43

[95] Kertesz M., Iovino N., Unnerstall U., Gaul U., and Segal E. (2007). The role of site accessibility in microRNA target recognition. *Nature Genetics*, 39(10):1278–1284. 7

[96] Ketting R.F., Fischer S.E.J., Bernstein E., Sijen T., Hannon G.J., et al. (2001). Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans. *Genes & Development*, 15(20):2654–2659. 3

[97] Kluiver J., Poppema S., de Jong D., Blokzijl T., Harms G., et al. (2005). BIC and miR-155 are highly expressed in Hodgkin, primary mediastinal and diffuse large B cell lymphomas. *The Journal of Pathology*, 207(2):243–249. 53

[98] Kong F., Sun C., Wang Z., Han L., Weng D., et al. (2011). miR-125b confers resistance of ovarian cancer cells to cisplatin by targeting pro-apoptotic Bcl-2 antagonist killer 1. *Journal of Huazhong University of Science and Technology. Medical Sciences = Hua Zhong Ke Ji Da Xue Xue Bao. Yi Xue Ying De Wen Ban = Huazhong Keji Daxue Xuebao. Yixue Yingdewen Ban*, 31(4):543–549. PMID: 21823019. 9

[99] Krek A., Grün D., Poy M.N., Wolf R., Rosenberg L., et al. (2005). Combinatorial microRNA target predictions. *Nature Genetics*, 37(5):495–500. PMID: 15806104. 63

[100] Krijgsman O., Roepman P., Zwart W., Carroll J.S., Tian S., et al. (2011). A diagnostic gene profile for molecular subtyping of breast cancer associated with treatment response. *Breast Cancer Research and Treatment*. PMID: 21814749. 13

[101] Küppers R., Schwering I., Bräuninger A., Rajewsky K., and Hansmann M.L. (2002). Biology of Hodgkin's lymphoma. *Annals of Oncology: Official Journal of the European Society for Medical Oncology / ESMO*, 13 Suppl 1:11–18. PMID: 12078890. 20

[102] Kushner P.J., Agard D.A., Greene G.L., Scanlan T.S., Shiau A.K., et al. (2000). Estrogen receptor pathways to AP-1. *The Journal of Steroid Biochemistry and Molecular Biology*, 74(5):311–317. PMID: 11162939. 14

[103] Lagos-Quintana M., Rauhut R., Lendeckel W., and Tuschl T. (2001). Identification of Novel Genes Coding for Small Expressed RNAs. *Science*, 294(5543):853–858. 3

[104] Lall S., Grün D., Krek A., Chen K., Wang Y., et al. (2006). A Genome-Wide Map of Conserved MicroRNA Targets in C. elegans. *Current Biology*, 16(5):460–471. 7

[105] Langmead B., Trapnell C., Pop M., and Salzberg S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, 10(3):R25. 42

[106] Lau N.C., Lim L.P., Weinstein E.G., and Bartel D.P. (2001). An Abundant Class of Tiny RNAs with Probable Regulatory Roles in Caenorhabditis elegans. *Science*, 294(5543):858–862. 3

[107] Lee R.C. and Ambros V. (2001). An Extensive Class of Small RNAs in Caenorhabditis elegans. *Science*, 294(5543):862–864. 3

[108] Lee R.C., Feinbaum R.L., and Ambros V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell*, 75(5):843–854. 3

[109] Lee Y.S. and Dutta A. (2007). The tumor suppressor microRNA let-7 represses the HMGA2 oncogene. *Genes & Development*, 21(9):1025–1030. 9

[110] Lenz G. and Staudt L.M. (2010). Aggressive lymphomas. *N Engl J Med*, 362(15):1417–29. Lenz, Georg Staudt, Louis M N Engl J Med. 2010 Apr 15;362(15):1417-29. 21, 84

[111] Lewis B.P., Burge C.B., and Bartel D.P. (2005). Conserved Seed Pairing, Often Flanked by Adenosines, Indicates that Thousands of Human Genes are MicroRNA Targets. *Cell*, 120(1):15–20. 57

[112] Lewis B.P., Shih I.h., Jones-Rhoades M.W., Bartel D.P., and Burge C.B. (2003). Prediction of Mammalian MicroRNA Targets. *Cell*, 115(7):787–798. 57

[113] Li B., Ruotti V., Stewart R.M., Thomson J.A., and Dewey C.N. (2010). RNA-Seq gene expression estimation with read mapping uncertainty. *Bioinformatics (Oxford, England)*, 26(4):493–500. PMID: 20022975. 42

[114] Li H. and Durbin R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*, 25(14):1754–1760. PMID: 19451168. 42

[115] Li H. and Durbin R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*, 26(5):589–595. PMID: 20080505. 42

[116] Li H., Ruan J., and Durbin R. (2008). Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Research*, 18(11):1851–1858. PMID: 18714091. 42

[117] Li Y., Zhang Y., Wang P., Chi Y., Zhang C., et al. (2012). Regression of A549 lung cancer tumors by anti-miR-150 vector. *Oncology Reports*, 27(1):129–134. PMID: 21935578. 9

[118] Liang P. and Pardee A.B. (1992). Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science (New York, N.Y.)*, 257(5072):967–971. PMID: 1354393. 23

[119] Lin C., Jackson A.L., Guo J., Linsley P.S., and Eisenman R.N. (2009). Myc-regulated microRNAs attenuate embryonic stem cell differentiation. *The EMBO Journal*, 28(20):3157–3170. 8

[120] Liu S., Guo W., Shi J., Li N., Yu X., et al. (2012). MicroRNA-135a contributes to the development of portal vein tumor thrombus by promoting metastasis in hepatocellular carcinoma. *Journal of Hepatology*, 56(2):389–396. 64

[121] Lu Y., Roy S., Nuovo G., Ramaswamy B., Miller T., et al. (2011). Anti-miR-222 and -181B suppresses growth of tamoxifen resistant xenografts in mouse by targeting TIMP3 and modulating mitogenic signal. *Journal of Biological Chemistry*. 9, 51, 54, 65

[122] Lykkesfeldt A.E. (1996). Mechanisms of tamoxifen resistance in the treatment of advanced breast cancer. *Acta Oncologica (Stockholm, Sweden)*, 35 Suppl 5:9–14. 53, 63

[123] Lykkesfeldt A.E. and Briand P. (1986). Indirect mechanism of oestradiol stimulation of cell proliferation of human breast cancer cell lines. *British Journal of Cancer*, 53(1):29–35. 54

[124] Lynch P.J. (????). *Breast anatomy normal scheme*. 12

[125] Lyng M., Laenkholm A., Sokilde R., Gravgaard K., Litman T., et al. (2012). Global microRNA expression profiling of high-risk estrogen receptor-positive breast cancers from patients receiving adjuvant Tamoxifen mono-therapy: a DBCG study. *Manuscript under review*. 65

[126] Massarweh S., Osborne C.K., Creighton C.J., Qin L., Tsimelzon A., et al. (2008). Tamoxifen Resistance in Breast Tumors Is Driven by Growth Factor Receptor Signaling with Repression of Classic Estrogen Receptor Genomic Function. *Cancer Research*, 68(3):826–833. 53

[127] Massarweh S. and Schiff R. (2006). Resistance to endocrine therapy in breast cancer: exploiting estrogen receptor/growth factor signaling crosstalk. *Endocrine-Related Cancer*, 13 Suppl 1:S15–24–S15–24. 53

[128] Mavrakis K.J., Wolfe A.L., Oricchio E., Palomero T., de Keersmaecker K., et al. (2010). Genome-wide RNA-mediated interference screen identifies miR-19 targets in Notch-induced T-cell acute lymphoblastic leukaemia. *Nature Cell Biology*, 12(4):372–379. PMID: 20190740. 8

[129] McCormick K.P., Willmann M.R., and Meyers B.C. (2011). Experimental design, preprocessing, normalization and differential expression analysis of small RNA sequencing experiments. *Silence*, 2(1):2. PMID: 21356093. 28

[130] McGee M. and Chen Z. (2006). Parameter Estimation for the Exponential-Normal Convolution Model for Background Correction of Affymetrix GeneChip Data. *Statistical Applications in Genetics and Molecular Biology*, 5(1). 33

[131] Mestdagh P., Van Vlierberghe P., De Weer A., Muth D., Westermann F., et al. (2009). A novel and universal method for microRNA RT-qPCR data normalization. *Genome Biology*, 10(6):R64. 36

[132] Miller T.E., Ghoshal K., Ramaswamy B., Roy S., Datta J., et al. (2008). MicroRNA-221/222 confers tamoxifen resistance in breast cancer by targeting p27Kip1. *The Journal of Biological Chemistry*, 283(44):29897–29903. 9, 51, 54

[133] Miller W.R. (2003). Aromatase inhibitors: mechanism of action and role in the treatment of breast cancer. *Seminars in Oncology*, 30(4 Suppl 14):3–11. PMID: 14513432. 15

[134] Millour J., Constantinidou D., Stavropoulou A.V., Wilson M.S.C., Myatt S.S., et al. (2010). FOXM1 is a transcriptional target of ERalpha and has a critical role in breast cancer endocrine sensitivity and resistance. *Oncogene*, 29(20):2983–2995. 63

[135] Miranda K.C., Huynh T., Tay Y., Ang Y., Tam W., et al. (2006). A Pattern-Based Method for the Identification of MicroRNA Binding Sites and Their Corresponding Heteroduplexes. *Cell*, 126(6):1203–1217. 7, 8

[136] Miranda K.C., Huynh T., Tay Y., Ang Y.S., Tam W.L., et al. (2006). A Pattern-Based Method for the Identification of MicroRNA Binding Sites and Their Corresponding Heteroduplexes. *Cell*, 126(6):1203–1217. 57

[137] Mortazavi A., Williams B.A., McCue K., Schaeffer L., and Wold B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods*, 5(7):621–628. PMID: 18516045. 42

[138] Mulder M.M.S., Heddema E.R., Pannekoek Y., Faridpooya K., Oud M.E.C.M., et al. (2006). No evidence for an association of ocular adnexal lymphoma with Chlamydia psittaci in a cohort of patients from the Netherlands. *Leukemia Research*, 30(10):1305–1307. 21

[139] Musgrove E.A. and Sutherland R.L. (2009). Biological determinants of endocrine resistance in breast cancer. *Nat Rev Cancer*, 9(9):631–643. 15, 16, 53

[140] O'Donnell K.A., Wentzel E.A., Zeller K.I., Dang C.V., and Mendell J.T. (2005). c-Myc-regulated microRNAs modulate E2F1 expression. *Nature*, 435(7043):839–843. 8

[141] Ota A., Tagawa H., Karnan S., Tsuzuki S., Karpas A., et al. (2004). Identification and Characterization of a Novel Gene, C13orf25, as a Target for 13q31-q32 Amplification in Malignant Lymphoma. *Cancer Research*, 64(9):3087–3095. 8

[142] Pandey D.P. and Picard D. (2009). miR-22 inhibits estrogen signaling by directly targeting the estrogen receptor alpha mRNA. *Molecular and Cellular Biology*, 29(13):3783–3790. 54

[143] R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. 31, 78

[144] Radhakrishnan V.M. and Martinez J.D. (2010). 14-3-3gamma induces oncogenic transformation by stimulating MAP kinase and PI3K signaling. *PLoS ONE*, 5(7):e11433. Radhakrishnan, Vijayababu M Martinez, Jesse D P30 CA023074/CA/NCI NIH HHS/United States R01 CA107510/CA/NCI NIH HHS/United States Research Support, N.I.H., Extramural United States PloS one PLoS One. 2010 Jul 2;5(7):e11433. 64

[145] Rai D., Karanti S., Jung I., Dahia P.L.M., and Aguiar R.C.T. (2008). Coordinated Expression of MicroRNA-155 and Predicted Target Genes in Diffuse Large B-cell Lymphoma. *Cancer genetics and cytogenetics*, 181(1):8–15. 53

[146] Ramsay D., Kent J., Hartmann R., and Hartman P. (2005). Anatomy of the lactating human breast redefined with ultrasound imaging. *Journal of Anatomy*, 206(6):525–534. PMID: 15960763 PMCID: 1571528. 11

[147] Reinartz J., Bruyns E., Lin J., Burcham T., Brenner S., et al. (2002). Massively parallel signature sequencing (MPSS) as a tool for in-depth quantitative gene expression profiling in all organisms. *Briefings in Functional Genomics & Proteomics*, 1(1):95–104. 28

[148] Reinhart B.J., Slack F.J., Basson M., Pasquinelli A.E., Bettinger J.C., et al. (2000). The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. *Nature*, 403(6772):901–906. 3

[149] Rhcastilhos (2012). retina. 19

[150] Riggins R.B., Schrecengost R.S., Guerrero M.S., and Bouton A.H. (2007). Pathways to tamoxifen resistance. *Cancer Letters*, 256(1):1–24. 16

[151] Ring A. and Dowsett M. (2004). Mechanisms of tamoxifen resistance. *Endocrine-Related Cancer*, 11(4):643–658. 53

[152] Ritchie M.E., Silver J., Oshlack A., Holmes M., Diyagama D., et al. (2007). A comparison of background correction methods for two-colour microarrays. *Bioinformatics*, 23(20):2700–2707. 78

[153] Robinson M.D., McCarthy D.J., and Smyth G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1):139–140. 43

[154] Rodriguez A., Griffiths-Jones S., Ashurst J.L., and Bradley A. (2004). Identification of Mammalian microRNA Host Genes and Transcription Units. *Genome Research*, 14(10a):1902–1910. 4

[155] Rosado M.F., Byrne J. Gerald E., Ding F., Fields K.A., Ruiz P., et al. (2006). Ocular adnexal lymphoma: a clinicopathologic study of a large cohort of patients with no evidence for an association with Chlamydia psittaci. *Blood*, 107(2):467–472. 21, 75

[156] Rose D.P. and Vona-Davis L. (2010). Interaction between menopausal status and obesity in affecting breast cancer risk. *Maturitas*, 66(1):33–38. PMID: 20181446. 12

[157] Rozowsky J., Euskirchen G., Auerbach R.K., Zhang Z.D., Gibson T., et al. (2009). PeakSeq enables systematic scoring of ChIP-seq experiments relative to controls. *Nature Biotechnology*, 27(1):66–75. PMID: 19122651. 42

[158] Ruijter J.M., Ramakers C., Hoogaars W.M.H., Karlen Y., Bakker O., et al. (2009). Amplification efficiency: linking baseline and bias in the analysis of quantitative PCR data. *Nucleic Acids Research*, 37(6):e45–e45. 27

[159] Sainsbury J.R.C., Anderson T.J., and Morgan D.A.L. (2000). Breast cancer. *BMJ : British Medical Journal*, 321(7263):745–750. PMID: 10999911 PMCID: 1127865. 13

[160] Saito Y., Liang G., Egger G., Friedman J.M., Chuang J.C., et al. (2006). Specific activation of microRNA-127 with downregulation of the proto-oncogene BCL6 by chromatin-modifying drugs in human cancer cells. *Cancer Cell*, 9(6):435–443. 75

[161] Sakhinia E., Glennie C., Hoyland J.A., Menasce L.P., Brady G., et al. (2007). Clinical quantitation of diagnostic and predictive gene expression levels in follicular and diffuse large B-cell lymphoma by RT-PCR gene expression profiling. *Blood*, 109(9):3922–3928. 85

[162] Sander S., Bullinger L., Klapproth K., Fiedler K., Kestler H.A., et al. (2008). MYC stimulates EZH2 expression by repression of its negative regulator miR-26a. *Blood*, 112(10):4202–12. Sander, Sandrine Bullinger, Lars Klapproth, Kay Fiedler, Katja Kestler, Hans A Barth, Thomas F E Moller, Peter Stilgenbauer, Stephan Pollack, Jonathan R Wirth, Thomas Blood. 2008 Nov 15;112(10):4202-12. Epub 2008 Aug 19. 82

[163] Sassen S., Miska E.A., and Caldas C. (2008). MicroRNA—implications for cancer. *Virchows Archiv*, 452(1):1–10. PMID: 18040713 PMCID: 2151131. 9, 75

[164] Schena M., Shalon D., Davis R.W., and Brown P.O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science (New York, N.Y.)*, 270(5235):467–470. PMID: 7569999. 23

[165] Schmieder R. and Edwards R. (2011). Quality control and preprocessing of metagenomic datasets. *Bioinformatics*, 27(6):863–864. PMID: 21278185 PMCID: 3051327. 41

[166] Schulte J.H., Horn S., Otto T., Samans B., Heukamp L.C., et al. (2008). MYCN regulates oncogenic MicroRNAs in neuroblastoma. *Int J Cancer*, 122(3):699–704. Schulte, Johannes H Horn, Sebastian Otto, Tobias Samans, Birgit Heukamp, Lukas C Eilers, Ursula-Christa Krause, Michael Astrahantseff, Kathy Klein-Hitpass, Ludger Buettner, Reinhard Schramm, Alexander Christiansen, Holger Eilers, Martin Eggert, Angelika Berwanger, Bernd Int J Cancer. 2008 Feb 1;122(3):699-704. 82

[167] Selbach M., Schwanhäusser B., Thierfelder N., Fang Z., Khanin R., et al. (2008). Widespread changes in protein synthesis induced by microRNAs. *Nature*, 455(7209):58–63. PMID: 18668040. 7

[168] Shields J.A., Shields C.L., and Scartozzi R. (2004). Survey of 1264 patients with orbital tumors and simulating lesions: The 2002 Montgomery Lecture, part 1. *Ophthalmology*, 111(5):997–1008. Shields, Jerry A Shields, Carol L Scartozzi, Richard Ophthalmology. 2004 May;111(5):997-1008. 75

[169] Sjö L.D., Heegaard S., Prause J.U., Petersen B.L., Pedersen S., et al. (2009). Extranodal marginal zone lymphoma in the ocular region: clinical, immunophenotypical, and cytogenetical characteristics. *Investigative Ophthalmology & Visual Science*, 50(2):516–522. 21, 75

[170] Smyth G.K. (2004). Linear Models and Empirical Bayes Methods for Assessing Differential Expression in Microarray Experiments. *Statistical Applications in Genetics and Molecular Biology*, 3. 33, 39, 55, 56, 78

[171] Smyth G.K. (2005). Limma: linear models for microarray data. In R. Gentleman, V. Carey, S. Dudoit, R. Irizarry, and W. Huber, editors, *Bioinformatics and Computational Biology Solutions using R and Bioconductor*, pages 397–420. Springer, New York. 33, 55, 56

[172] Smyth G.K. (2005). Limma: linear models for microarray data. In R. Gentleman, V. Carey, S. Dudoit, R. Irizarry, and W. Huber, editors, *Bioinformatics and Computational Biology Solutions using R and Bioconductor*, pages 397–420. Springer, New York. 78

[173] Smyth G.K. and Speed T. (2003). Normalization of cDNA microarray data. *Methods (San Diego, Calif.)*, 31(4):265–273. 78

[174] Song M., Lee K., and Kang D. (2011). Breast cancer prevention based on gene-environment interaction. *Molecular Carcinogenesis*, 50(4):280–290. PMID: 21465576. 12

[175] Sørlie T., Perou C.M., Tibshirani R., Aas T., Geisler S., et al. (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proceedings of the National Academy of Sciences of the United States of America*, 98(19):10869–10874. PMID: 11553815. 13

[176] Sotiriou C., Neo S., McShane L.M., Korn E.L., Long P.M., et al. (2003). Breast cancer classification and prognosis based on gene expression profiles from a population-based study. *Proceedings of the National Academy of Sciences of the United States of America*, 100(18):10393–10398. PMID: 12917485. 13

[177] Spicer D.V. and Pike M.C. (2000). Future possibilities in the prevention of breast cancer: Luteinizing hormone-releasing hormone agonists. *Breast Cancer Research*, 2(4):264–267. PMID: 11250719 PMCID: 138786. 15

[178] Streit S., Michalski C.W., Erkan M., Kleeff J., and Friess H. (2008). Northern blot analysis for detection and quantification of RNA in pancreatic cancer cells and tissues. *Nat. Protocols*, 4(1):37–43. 26

[179] Streubel B., Lamprecht A., Dierlamm J., Cerroni L., Stolte M., et al. (2003). T(14;18)(q32;q21) involving IGH and MALT1 is a frequent chromosomal aberration in MALT lymphoma. *Blood*, 101(6):2335–2339. PMID: 12406890. 21

[180] Su H., Trombly M.I., Chen J., and Wang X. (2009). Essential and overlapping functions for mammalian Argonautes in microRNA silencing. *Genes & Development*, 23(3):304–317. PMID: 19174539. 6

[181] Swerdlow S.H.e.a. (2008). *WHO classification of tumours of haematopoietic and lymphoid tissues*. International Agency for Research on Cancer, Lyon, France. 20, 21

[182] Vandesompele J., De Preter K., Pattyn F., Poppe B., Van Roy N., et al. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biology*, 3(7):research0034. 36

[183] Vargas R.L., Fallone E., Felgar R.E., Friedberg J.W., Arbini A.A., et al. (2006). Is there an association between ocular adnexal lymphoma and infection with Chlamydia psittaci? The University of Rochester experience. *Leukemia Research*, 30(5):547–551. 21

[184] Velculescu V.E., Zhang L., Vogelstein B., and Kinzler K.W. (1995). Serial analysis of gene expression. *Science (New York, N.Y.)*, 270(5235):484–487. PMID: 7570003. 23, 28

[185] Wang H., Garzon R., Sun H., Ladner K.J., Singh R., et al. (2008). NF-kappaB-YY1-miR-29 regulatory circuitry in skeletal myogenesis and rhabdomyosarcoma. *Cancer Cell*, 14(5):369–381. 83, 85

[186] Wang J., Lu M., Qiu C., and Cui Q. (2009). TransmiR: a transcription factor-microRNA regulation database. *Nucleic Acids Research*, 38(Database):D119–D122–D119–D122. 78, 82

[187] Wang X. (2008). miRDB: a microRNA target prediction and functional annotation database with a wiki interface. *RNA (New York, N.Y.)*, 14(6):1012–1017. 57

[188] Wilcoxon F. (1945). Individual Comparisons by Ranking Methods. *Biometrics Bulletin*, 1(6):80–83. 39

[189] Woo J., Tang C., Rho M., Lee J., Kwon H., et al. (2006). The Clinical Characteristics and Treatment Results of Ocular Adnexal Lymphoma. *Korean Journal of Ophthalmology : KJO*, 20(1):7–12. 20, 75

[190] Workman C., Jensen L.J., Jarmer H., Berka R., Gautier L., et al. (2002). A new non-linear normalization method for reducing variability in DNA microarray experiments. *Genome Biology*, 3(9):research0048.1–research0048.16–research0048.1–research0048.16. 34, 37, 38, 55

[191] Wyman S.K., Knouf E.C., Parkin R.K., Fritz B.R., Lin D.W., et al. (2011). Post-transcriptional generation of miRNA variants by multiple nucleotidyl transferases contributes to miRNA transcriptome complexity. *Genome Research*, 21(9):1450–1461. PMID: 21813625. 42

[192] Xiao F., Zuo Z., Cai G., Kang S., Gao X., et al. (2009). miRecords: an integrated resource for microRNA–target interactions. *Nucleic Acids Research*, 37(Database issue):D105–D110. PMID: 18996891 PMCID: 2686554. 8

[193] Yamakuchi M., Lotterman C.D., Bao C., Hruban R.H., Karim B., et al. (2010). P53-induced microRNA-107 inhibits HIF-1 and tumor angiogenesis. *Proceedings of the National Academy of Sciences of the United States of America*, 107(14):6334–6339. PMID: 20308559. 8

[194] Yang H., Kong W., He L., Zhao J.J., O'Donnell J.D., et al. (2008). MicroRNA expression profiling in human ovarian cancer: miR-214 induces cell survival and cisplatin resistance by targeting PTEN. *Cancer Research*, 68(2):425–433. 9, 53

[195] Yang X., Feng M., Jiang X., Wu Z., Li Z., et al. (2009). miR-449a and miR-449b are direct transcriptional targets of E2F1 and negatively regulate pRb–E2F1 activity through a feedback loop by targeting CDK6 and CDC25A. *Genes & Development*, 23(20):2388–2393. 63

[196] Ye G., Fu G., Cui S., Zhao S., Bernaudo S., et al. (2011). MicroRNA 376c enhances ovarian cancer cell survival by targeting activin receptor-like kinase 7: implications for chemoresistance. *Journal of Cell Science*, 124(3):359–368. 53

[197] Zhang X., Chen X., Lin J., Lwin T., Wright G., et al. (2011). Myc represses miR-15a|[sol]|miR-16-1 expression through recruitment of HDAC3 in mantle cell and other non-Hodgkin B-cell lymphomas. *Oncogene*. 85

[198] Zhao F., Xuan Z., Liu L., and Zhang M.Q. (2005). TRED: a Transcriptional Regulatory Element Database and a platform for in silico gene regulation studies. *Nucleic Acids Research*, 33(Database Issue):D103–D107. PMID: 15608156 PMCID: 539958. 14

[199] Zhao J., Lin J., Yang H., Kong W., He L., et al. (2008). MicroRNA-221/222 Negatively Regulates Estrogen Receptor and Is Associated with Tamoxifen Resistance in Breast Cancer. *Journal of Biological Chemistry*, 283(45):31079 –31086. 51, 54

[200] Ziegelbauer J.M., Sullivan C.S., and Ganem D. (2009). Tandem array-based expression screens identify host mRNA targets of virus-encoded microRNAs. *Nature Genetics*, 41(1):130–134. PMID: 19098914. 7